REPORT NUMBER 21

# QUICKEST DETECTION PROCEDURES AND TRANSIENT SIGNAL DETECTION

BRUCE BRODER AND STUART C. SCHWARTZ

INFORMATION SCIENCES AND SYSTEMS LABORATORY

Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544



NOVEMBER 1990

Prepared for

OFFICE OF NAVAL RESEARCH (Code 1111)
Statistics and Probability Branch
Arlington, Virginia 22217
under Contract N00014-88-K0162
Acoustic Transient Initiative

S.C. Schwartz, Principal Investigator

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>21 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Quickest Detection Procedures and Transient Signal Detection | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>Sept. 1988–June 1990<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Bruce Broder and Stuart Schwartz | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-81-K-0162 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Information Sciences & Systems Laboratory<br>Dept. of Electrical Engineering<br>Princeton University, Princeton, NJ 08544 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>411m006 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research (Code 1111SP)<br>Department of the Navy<br>Arlington, VA 22217 | | 12. REPORT DATE<br>November 1990<br>13. NUMBER OF PAGES<br>192 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Also submitted as the Ph.D. thesis of Bruce Broder to the EE Department, Princeton University, Princeton, NJ 08544, June 1990

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

quickest detectors
transient detection
Page's test
disorder problem
Cusum test
Gabor representation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This dissertation focuses on sequential decision procedures to detect changes in the statistical model of an observed random process when these changes can occur at unknown times. In the disorder problem, the samples are drawn according to one statistical model until some unknown time after which all the samples correspond to a second model. In the transient problem, the model then reverts back to the first one after some finite time.

Continued

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 68 IS OBSOLETE
S/N 0102-LF-014-6601

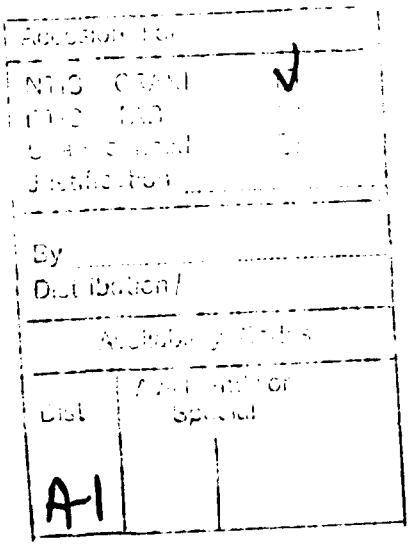
Continued

A detection procedure known as Page's test is investigated for the quick detection of the disorder. A simple asymptotic measure is defined and an analytic formula is developed which can be useful in evaluating the performance of Page's test in various situations. By examining the local performance, it is found that the performance is directly related to the efficacy in the binary hypothesis testing situation, allowing the wealth of results in that context to be transferred to Page's test.

Because Page's test appears to detect changes in distribution quickly, it is a candidate for transient signal detection. It is used to detect transients of arbitrary spectral shape by developing the test based on the DFT of the observations. This test is compared against other sequential detection schemes and it is shown that Page's test performs very favorably and is easier to implement.

The Gabor signal representation in which a signal is modeled by exponentially decaying sinusoids is studied for its use in transient signal detection. A detection algorithm is develcped using the coefficients of these basis functions. The focus is on the procedure for computing the coefficients and the effect on detection performance. A sequential algorithm for computing the coefficients is derived so that a real-time detection scheme can be implemented.

Finally, the optimal sample size for the detection of a transient signal which can occur any time during a fixed observation period is determined. The optimization is done assuming either that the transient can occur with equal probability any time during the integration period or that it occurs at the worst possible time during the period.

# Quickest Detection Procedures and Transient Signal Detection

*Bruce Broder*

## Abstract

This dissertation focuses on sequential decision procedures to detect changes in the statistical model of an observed random process when these changes can occur at unknown times. In the disorder problem, the samples are drawn according to one statistical model until some unknown time after which all the samples correspond to a second model. In the transient problem, the model then reverts back to the first one after some finite time.

A detection procedure known as Page's test is investigated for the quick detection of the disorder. A simple asymptotic measure is defined and an analytic formula is developed which can be useful in evaluating the performance of Page's test in various situations. By examining the local performance, it is found that the performance is directly related to the efficacy in the binary hypothesis testing situation, allowing the wealth of results in that context to be transferred to Page's test.

Because Page's test appears to detect changes in distribution quickly, it is a candidate for transient signal detection. It is used to detect transients of arbitrary spectral shape by developing the test based on the DFT of the observations. This test is compared against other sequential detection schemes and it is shown that Page's test performs very favorably and is easier to implement.

The Gabor signal representation in which a signal is modeled by exponentially decay-

ing sinusoids is studied for its use in transient signal detection. A detection algorithm is developed using the coefficients of these basis functions. The focus is on the procedure for computing the coefficients and the effect on detection performance. A sequential algorithm for computing the coefficients is derived so that a real-time detection scheme can be implemented.

Finally, the optimal sample size for the detection of a transient signal which can occur any time during a fixed observation period is determined. The optimization is done assuming either that the transient can occur with equal probability any time during the integration period or that it occurs at the worst possible time during the period.

# Table of Contents

# — 1 —

# Introduction

## 1.1 Motivation

This dissertation focuses on sequential decision procedures to detect changes in the statistical model of an observed random process when these changes can occur at unknown times. In classical sequential detection procedures such as those considered by Wald [9], it is assumed that all the time samples come from one statistical model or from an alternative, i.e., the sequential binary hypothesis situation. The problem of detecting the change from one statistical model to a second model had not been considered until the works of Page [3] and Shiryaev [5–8].

Two types of detection problems which depend on the time element are considered in this work: the disorder problem and the transient problem. In the disorder problem, the samples are drawn according to one statistical model until some unknown time after which all the samples correspond to a second model. In the quickest detection formulation of this problem, it is desired to detect the time of change with minimum delay for a fixed mean time between false alarms. This is in contrast to the approach in the classical detection problem

where the probability of detection is maximized for a fixed probability of false alarm. The quickest detection problem was first considered by Page [3] in the context of quality control in a manufacturing process where it is desired to detect when the machinery fails and the items subsequently produced are defective. This model also has applications in the area of signal processing. For example, it could model a sonar observation of a submarine which suddenly turns on its engines. In the transient problem, the samples again come from one model until an unknown time, after which the samples are drawn from a second model which lasts for a known or unknown but finite number of samples. Then, the samples are again drawn according to the first statistical model. In this case, the performance of a sequential detection procedure is evaluated according to the probability of detecting the transient and the mean time between false alarms. The transient model is also applicable to a variety of signal processing problems.

## 1.2 Overview

The body of this dissertation is divided into four chapters as summarized below. With the exception of Chapter 3, which uses some material from Chapter 2, each of the chapters can be read and understood independently.

In Chapter 2, a class of detection procedures is investigated in the context of non-parametric quickest detection. The procedures are a generalization of a test first studied by Page in 1954 [3] to detect a change in the distribution of random variables observed sequentially in time. The goal is to detect the unknown time of change with minimum delay while insuring infrequent false alarms. Originally, Page's test was implemented with

the log-likelihood ratio for the two distributions under consideration. In this work, the test is generalized by implementing it with other nonlinearities which do not depend on the distributions. The linear detector, sign detector and dead zone limiter are investigated in detail, but the theoretical results apply to arbitrary nonlinearities.

Methods for generating performance curves for Page's test are discussed. Based on the performance curves, a simple asymptotic measure is defined and an analytic formula is developed which can be useful in evaluating the performance of Page's test using different nonlinearities for various noise distributions. Also, by examining the local performance of Page's test, it was found that the performance is directly related to the efficacy in the classical binary hypothesis testing situation. Thus, the wealth of results on using memoryless nonlinearities in classical detection theory can be transferred to the quickest detection problem using Page's test.

Because Page's test detects changes in distribution quickly, one might expect that it would be useful for detecting transient signals which occur at unknown times. In Chapter 3, the performance of Page's test is evaluated for transient signal detection using the probability of detection and the mean time between false alarms as the performance criteria. By developing a test based on the Discrete Fourier Transform of the observations, Page's test can be used to detect transients of arbitrary spectral shape. This test is compared against other sequential detection schemes including one proposed by Wolcin [10]. The results indicate that Page's test performs as well as the other tests in almost all the examples considered and it is also the easiest to implement.

In Chapter 4, a signal representation due to Gabor [2] is studied because of its potential use for transient signal detection. This idea was first suggested by Friedlander and Porat

[1]. Essentially, the transient is modeled by a group of exponentially decaying sinusoids with arbitrary arrival times. By decomposing a signal into a set of such basis functions and finding the corresponding coefficients, a detection algorithm is developed. This chapter focuses on the two methods of computing the coefficients, the biorthogonal function method used in [1] and maximum likelihood estimation. It is demonstrated that detectors based on the maximum likelihood estimates outperform those based on the other method. A sequential algorithm for computing the maximum likelihood estimates is derived so that a real-time detection scheme can be implemented.

Often when a real-time signal processor is built in hardware, the incoming data is buffered into fixed-size blocks to facilitate the implementation. In Chapter 5, the optimal block size for transient signal detection is determined. This is motivated by the work of Pelkowitz and Schwartz [4] in determining the optimal block size for the quickest detection problem. For the present situation, the sample size is optimized so as to maximize the probability of detecting a transient that can occur at any time during the integration period. The two optimization criteria considered are similar to those found in [4], the mean criterion and the minimax criterion. In the mean criterion, the sample size is chosen to maximize the mean probability of detection, assuming that the transient is equally likely to arrive any time during the integration period. In the minimax criterion, the optimal sample size maximizes the probability of detection for the worst possible arrival time.

Finally, the original results of this dissertation are summarized in Chapter 6 along with suggested areas for further research.

## 1.3 References


[1] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-37, no. 2, pp. 169–180, February 1989.

[2] D. Gabor, "Theory of Communication," *Proc. of the IEEE*, vol. 93, pt. III, pp. 429–457, 1946.

[3] E. S. Page, "Continuous Inspection Schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[4] L. Pelkowitz and S. C. Schwartz, "Asymptotically optimum sample size for quickest detection," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-23, no. 2, pp. 263–272, March 1987.

[5] A. N. Shiryaev, "The detection of spontaneous effects," *Sov. Math. Dokl.*, no. 2, pp. 740–743, 1961.

[6] ——, "The problem of the most rapid detection of a disturbance in a stationary regime," *Sov. Math. Dokl.*, no. 2, pp. 795–799, 1961.

[7] ——, "On optimum methods in quickest detection problems." *Theory Prob. Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[8] ——, "Some exact formulas in a disorder problem," *Theory Prob. Appl.*, vol. 10, no. 3, pp. 348–354, 1965.

[9] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.

[10] J. J. Wolcin, "Maximum likelihood detection of transient signals using sequenced short-time power spectra," TM 831138, Naval Underwater Systems Center, August 1983.

# — 2 —

# Nonparametric Quickest Detection and The Performance of Page's Test

## 2.1 Introduction

When a change in the probability laws of a random process occurs, a *disorder* is said to take place. The time at which the disorder occurs is called the *disorder time* and is either an unknown parameter or a random variable with a known, assumed, or unknown distribution. In the simplest case, a discrete-time random process, $X_i, i = 1,2,3,\ldots$ is considered [36,37]. Before the disorder time, $m$, the random variables $X_1, X_2, \ldots, X_{m-1}$ are independent and identically distributed (i.i.d.) with distribution function $F_0$, and after the disorder, the random variables $X_m, X_{m+1}, \ldots$ are i.i.d. with distribution $F_1$ and are independent of $X_1, X_2, \ldots, X_{m-1}$. This model was originally used for quality control in the manufacturing of a product where it is desired to detect when the machinery fails and the items subsequently produced are defective [31]. The same model, however, can also be used for signals in a radar or sonar observation. For example, with continuous-time observations, consider a Brownian motion, $B_t$, whose drift changes at the disorder time $t_0$

[37–39]. Specifically, let $B_t$ satisfy the stochastic differential equation

$$dB_t = u(t - t_0)dt + dw_t$$

where $w_t$ is a Wiener process and $u(t)$ is the unit step function

$$u(t) = \begin{cases} 1, & \text{for } t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

This might be the situation when looking at a radar return with no targets present, when, at some unknown time, a target comes over the horizon into the view of the radar.

Detecting the disorder can be thought of as a classical decision problem with an infinite number of hypotheses, one hypothesis for each possible disorder time. In this sense, detecting the disorder is more akin to an estimation problem where a combined "detector–estimator" can either declare that a disorder is not present or give an estimate of the disorder time when declaring the disorder present.

There are two distinct algorithmic approaches to the disorder problem, *on–line* and *off–line* processing [2]. In the latter, the random process is observed over a finite time interval $[t_1, t_2]$ or only a finite number of samples are used [3,7,17,32,34,35,40]. The problem is then one of classical hypothesis testing,

$$H_0 : \text{No disorder occurs in } [t_1, t_2]$$

*versus*

$$H_1 : \text{A disorder occurs in } [t_1, t_2].$$

In this case, the performance is measured in terms of the probability of detection versus the probability of false alarm. Often, in order to obtain tractable results, the number of samples is made asymptotically infinite [13–16]. In on-line processing, the observations are

processed sequentially and a finite observation interval is not assumed. Here, performance is measured in terms of delay in detection when the disorder occurs versus the mean time between false alarms when the disorder is absent. When this approach is adopted, the disorder problem is termed the "quickest detection" problem to indicated the performance criteria used for the sequential processing. This approach is taken in [1,4,21,25,31,36–39] and will be used throughout this chapter.

There are two probabilistic approaches to looking at the disorder problem, the Bayesian and non–Bayesian or maximum likelihood approach. In the Bayesian approach [7,40], which was first studied extensively by Shiryaev [36,37], a prior distribution is assumed for the disorder time and one attempts to minimize some cost functional. Often the geometric and exponential distributions are assumed for the discrete and continuous times cases, respectively.

In the non–Bayesian or maximum likelihood approach to the disorder problem, no prior distribution is assumed for the disorder time. In off-line processing this is equivalent to assuming a uniform prior distribution over the observation interval, resulting in a detector which computes the likelihood function of the data for all possible disorder times. In on-line processing, the disorder time is considered an unknown parameter.

This work will consider only the non–Bayesian formulation of the quickest detection problem. The situation is as follows. Consider the sequential observation of independent random variables $X_1, X_2, \ldots$ such that $X_1, \ldots, X_{m-1}$ have distribution function $F_0$ while $X_m, X_{m+1}, \ldots$ have distribution function $F_1 \neq F_0$. The two distributions may not be known exactly and the disorder time $m$ is unknown. In this setting, it is necessary to define appropriate performance criteria which do not depend on the true disorder time.

As formulated by Lorden [25] for the discrete–time case, the criteria we use are the mean time between false alarms and the expected delay which are defined as follows. Let $N$ be the stopping variable associated with the sequential detection algorithm. Let $P_m$ denote the probability measure when the disorder time is $m$ and let $E_m$ denote the corresponding expectation. Similarly, let $P_0$ and $E_0$ correspond to the case in which $X_1, X_2, \ldots$ all have distribution $F_0$. Define the worst case expected delay in detection $D$ according to

$$D = \sup_{m \geq 1} \operatorname{ess\,sup} E_m\{\max(0, N - m + 1) | \mathcal{F}_{m-1}\},$$

where $\mathcal{F}_m$ is the sigma algebra generated by $\{X_1, \ldots, X_m\}$. In other words, $D$ is the smallest value such that for every $m = 1, 2, \ldots$

$$\mathrm{E}_m\{\max(0, N - m + 1) | \mathcal{F}_{m-1}\} \leq D$$

almost surely under $P_0$. The goal of the quick detection of the change is in competition with the desire for a large mean time between false alarms $T$ which is defined by $T = E_0(N)$. Thus, the performance of a given algorithm is specified by the pair $(T, D)$. According to these criteria, the optimal detector would have the minimum expected delay $D$ for a fixed mean time between false alarms $T$.

Using Lorden's criteria for the discrete–time case of i.i.d. observations and a change in distribution, we will investigate one class of detectors which is sufficiently rich to include the optimal detector. In particular, we look at a generalization of a control chart procedure for a manufacturing process which was studied extensively by Page in 1954 [31]. Define a cumulative sum statistic $S_n$ according to

$$\begin{aligned} S_n &= \sum_{i=1}^{n} g(X_i) \\ S_0 &= z. \end{aligned} \tag{1}$$

Here, $g(\bullet)$ is some function and $z$ is a parameter called the *initial score*. For a given function $g$ and threshold $h$, let the stopping variable $N$ be defined as

$$N = \inf\{n \geq 1 : S_n > h\}. \tag{2}$$

The test procedure given by (1) and (2) has been studied only when $g(\bullet)$ is the log–likelihood ratio, $g(x) = \log[dF(x|\theta_1)/dF(x|\theta_0)]$. For this case, the detector has been given various names in the literature including Page's test, the cum-sum, cusum, or cumulative sum test, and Hinkley's detector. We will refer to the test with arbitrary nonlinearities also as Page's test.

When Page's test is implemented with the log–likelihood ratio, it is equivalent to a sequential implementation of the maximum likelihood detector. The log–likelihood function of the observations $l(X_1, \ldots, X_n)$ is

$$l(X_1, \ldots, X_n) = \sum_{i=1}^{m-1} \log[dF_0(X_i)] + \sum_{i=m}^{n} \log[dF_1(X_i)]$$

which can be rewritten

$$l(X_1, \ldots, X_n) = \sum_{i=m}^{n} \log\left[\frac{dF_1(X_i)}{dF_0(X_i)}\right] + \sum_{i=1}^{n} \log\left[dF_0(X_i)\right].$$

Since the last term does not depend on the disorder time, the maximum of the likelihood function is obtained by using (1)-(2) with the log-likelihood ratio and initial score $S_0 = 0$. Lorden proved that for $T \geq T_0$, this detector minimizes the delay in the class of all detectors of form (1)-(2), asymptotically as $T_0$ tends to infinity. In [29], Moustakides extended this result to the non–asymptotic case.

Dyson [9] investigated the locally optimal or weak signal version of Lorden's result. Specifically, let $P_0$ be the distribution of the i.i.d. observations before the disorder and let $P_\theta$ be the distribution of the i.i.d. observations after the disorder. We assume that $\theta$

parameterizes a family of distributions $\{P_\theta,\ \theta \geq 0\}$ which includes $P_0$ and $\lim_{\theta \to 0} P_\theta = P_0$. Let $p_\theta(x)$ be the probability density function corresponding to $P_\theta$ which is assumed to exist and to be differentiable with respect to $\theta$. Then, for small $\theta$ the optimal detector in the sense of Lorden is Page's test with $g(x)$ given by

$$g(x) = \theta \left\{ \frac{\partial p_\theta(x)}{\partial \theta} \frac{1}{p_\theta(x)} \right\}_{\theta=0} - B(\theta)$$

where $B(\theta)$ is of $O(\theta^2)$.

In this chapter we investigate the performance of Page's test for arbitrary nonlinearities $g(x)$ in the local and non-local situations. In particular, we will look at nonlinearities which do not require knowledge of the probability distributions before and after the disorder. In that sense, the results have applications in nonparametric quickest detection situations. Specifically, we will examine Page's test when implemented with the linear detector, the sign detector, or the dead zone limiter. The analysis for the linear detector can be applied to a large class of nonlinearities by using a transformation of random variables. The sign detector and dead zone limiter are a special cases of a class of detectors for which the performance can be obtained in closed form. It should be noted, however, that the performance of Page's test with a particular nonlinearity will depend on the actual distributions of the observations so that the detector is not nonparametric in the most strict sense.

All previous work in nonparametric detection in the disorder problem has been done in the off-line setting. In [33], Page uses a fixed sample size (off-line) version of (1)-(2) implementing the sign detector and evaluates the performance in several examples. Tests based on the ranks and signs of the data are examined in [3,17]. Other nonparametric tests include extending the classical tests such as the Kolmogorov-Smirnov test and the Cramer-Von Mises test to the disorder problem [8]. These tests look for a change in the empirical cumulative distribution function to determine that a disorder has occurred and

their performance does not depend on the true distributions. Also in an off–line framework, Pettitt [34,35] uses an extended Mann-Whitney test which uses only the signs of the differences of the observations and is also truly nonparametric. The only related work in the sequential framework is the analysis of Page's test with observations drawn from $\{-1,+1\}$ done by Basseville [1], but the results were not applied in the nonparametric framework.

Because (1) involves sums of a memoryless nonlinear function of the observations, one might expect there to be a relationship to the classical detection problem

$$H_0: \quad X_1, X_2, \ldots, X_n \text{ are i.i.d. } F_0 \text{ distributed}$$

*versus*

$$H_1: \quad X_1, X_2, \ldots, X_n \text{ are i.i.d. } F_1 \text{ distributed}$$

using $S_n$ as the test statistic. For the case when $F_1(x) = F_0(x-\theta)$ corresponding to a shift in the mean of the distribution $F_0$, the *efficacy* $\mathcal{E}$ of the detector is a measure of the detector performance in the classical problem for a vanishingly small shift in mean and asymptotically large sample size. In Section 5, it will be shown that the performance of Page's test in the weak signal case is related to the detector efficacy. For the memoryless nonlinearity $g(\bullet)$ the efficacy is

$$\mathcal{E} = \lim_{n\to\infty} \lim_{\theta\to 0} \frac{\left(\frac{d}{d\theta} E[S_n|H_1]\right)^2}{n \operatorname{Var}[S_n|H_0]}$$

$$= \frac{(E[g'(X)|H_0])^2}{\operatorname{Var}[g(X)|H_0]}.$$

Larger values of the efficacy indicate better detector performance. The nonlinearity which maximizes the efficacy, the *locally optimal nonlinearity* $g_{lo}(x)$, can be shown to be [19,23,28]

$$g_{lo}(x) = -\frac{f_0'(x)}{f_0(x)}$$

where $f_0$ is the probability density function for $F_0$. In this case, the efficacy is equal to Fisher's Information $\mathcal{I}_f$ [22]

$$\mathcal{E} = \mathcal{I}_f = \int_{-\infty}^{\infty} f_0(x) \left[ \frac{f_0'(x)}{f_0(x)} \right]^2 dx.$$

The remainder of this chapter is organized as follows. In Section 2, a method is presented to compute performance curves for Page's test and several examples are given. The performance curves suggest an asymptotic performance measure which is defined and analyzed in Section 3. Properties of this measure and examples of its use are presented in Section 4. The performance of Page's test in the weak signal situation is analyzed in Section 5 and conclusions are presented in Section 6.

## 2.2 Performance Curves

We have seen that the mean time between false alarms is $T = E_0(N)$. Although not true in general, the expected delay given by (1) is simply $D = E_1(N)$ for Page's test [2.25]. Thus, the computation of both of these quantities, $T$ and $D$, simplifies to determining the average sample number (ASN) of Page's test when all the observations are independent and identically distributed with distribution $F$, that is when no change in distribution occurs. Several approaches have been taken to compute the ASN for Page's test. Some have used Monte Carlo simulations as in [1], but this will be computationally expensive if results are desired for a large mean time between false alarms. In [6], the interval between 0 and $h$ is discretized and the resulting random sequence is analyzed using Markov chains. For fixed values of $h$, this approach seems promising. However, in order to generate performance

curves, pairs of $(T, D)$ are required for many values of $h$. The Markov chain approach requires the inversion of a matrix which does not permit a recursive implementation as $h$ increases. Our approach involves discretizing the integral equations which are required to compute the ASN and solving a system of linear equations. This approach has been used in [12] and we will propose an alternate technique which allows recursive computation as $h$ increases.

To determine the performance curves indicating expected delay versus mean time between false alarms for Page's test, it is useful to reformulate the sequential procedure in (1)–(2). It is not difficult to see that the following test is equivalent [31]. Define the test statistic $\tilde{S}_n$ recursively by

$$\begin{aligned} \tilde{S}_n &= \max\{0, \tilde{S}_{n-1} + g(X_n)\} \\ \tilde{S}_0 &= z, \end{aligned} \tag{3}$$

and the stopping variable $N$ is given by

$$N = \inf\{n : \tilde{S}_n > h\}. \tag{4}$$

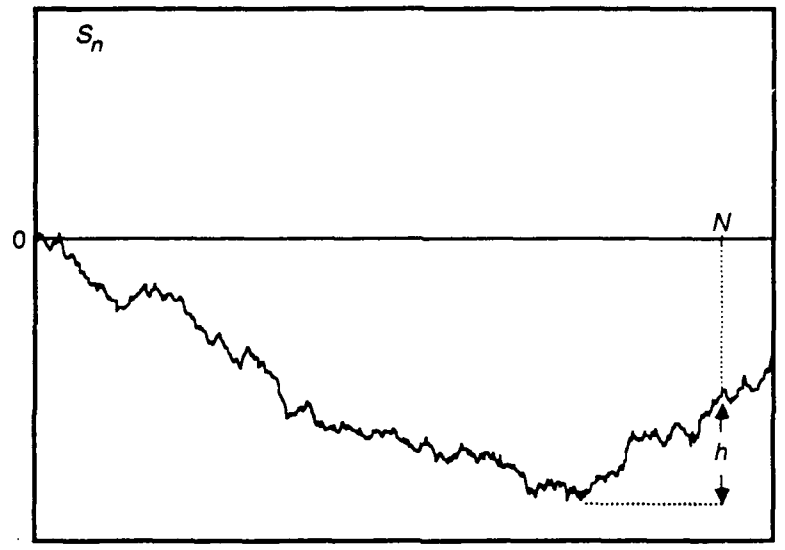
This formulation can be thought of as a repeated Wald sequential test with boundaries 0 and $h$ in the following sense. Let $M$ be the stopping variable for the Wald test given by

$$M = \inf\{n : S_n < 0 \text{ or } S_n > h\} \tag{5}$$

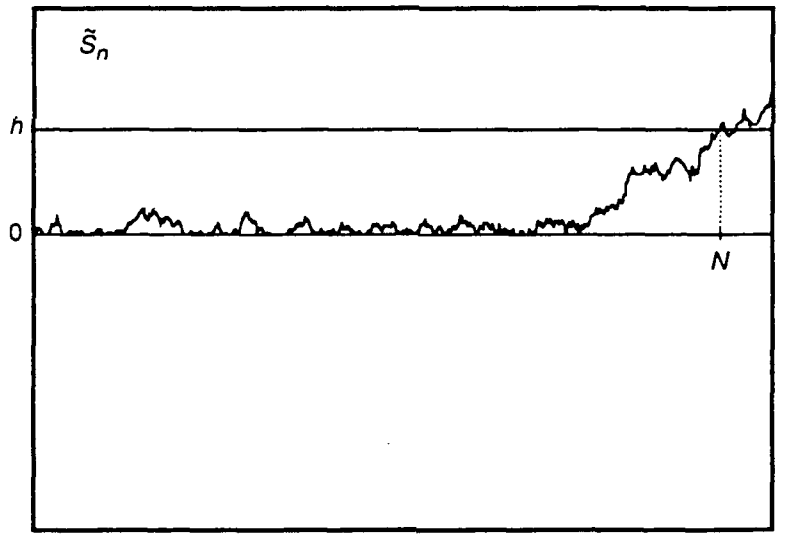
where $S_n$ is defined in (1). If the Wald test hits the lower boundary 0 first, then another Wald test is performed, and so on, until such time that a Wald test exceeds the threshold $h$. See Figure 1 for an illustration of the equivalent implementations of Page's test in (2) and (4).

For convenience, we parameterize the distributions as follows,

$$F(x|\theta) = \begin{cases} F_0(x), & \text{if } \theta = \theta_0 \\ F_1(x), & \text{if } \theta = \theta_1 \end{cases}$$




(a)




(b)

**Figure 1.** Two equivalent implementations of Page's test. In (a), the cumulative sum statistic $S_n$ is compared to the previous minimum value of the statistic. In (b), a two–sided sequential test is implemented in which the test statistic $\tilde{S}_n$ is reset to zero if it would have fallen below zero.

where $\theta_0$ and $\theta_1$ are some parameters such as the means of the random variables before and after the disorder. Denote by $\mathcal{N}(z,\theta)$ the average sample number of test (4) with initial score $\tilde{S}_0 = z$ when the $\{X_i\}$ are i.i.d. $F(\theta)$ distributed, i.e.,

$$\mathcal{N}(z,\theta) = E[N|S_0 = z,\theta].$$

Similarly, let $\mathcal{M}(z,\theta)$ be the ASN of the Wald test (5) and let $\mathcal{P}(z,\theta)$ be the operating characteristic (OC) of the same Wald test, that is, the probability of hitting the lower boundary first

$$\mathcal{P}(z,\theta) = P\{S_M < 0\}.$$

Then, the ASN of Page's test can be computed from [31]

$$\mathcal{N}(z,\theta) = \frac{\mathcal{M}(0,\theta)}{1-\mathcal{P}(0,\theta)}\mathcal{P}(z,\theta) + \mathcal{M}(z,\theta).$$

The mean time between false alarms $T$ is the average sample number of test (4) with initial score $\tilde{S}_0 = 0$ and $X_i \sim F(\theta_0)$,

$$T = \mathcal{N}(0,\theta_0) = \frac{\mathcal{M}(0,\theta_0)}{1-\mathcal{P}(0,\theta_0)} \tag{6}$$

Recall that the delay in detection $D$ as defined by Lorden is in effect the worst possible expected delay. Using the formulation in (4) (see Figure 1(b)), it is obvious that the worst case delay occurs when the test statistic is zero at the disorder time, $\tilde{S}_m = 0$, since then the test statistic has the furthest to travel when the disorder occurs. A proof of this result can be found in [2,25]. Thus, the delay is

$$D = \mathcal{N}(0,\theta_1) = \frac{\mathcal{M}(0,\theta_1)}{1-\mathcal{P}(0,\theta_1)}. \tag{7}$$

In this way, calculation of the performance curves of $D$ *vs.* $T$ only requires calculating the ASN and OC of a Wald test with initial score $\tilde{S}_0 = 0$ for the two possible distributions.

When the linearity $g(x) = x$ is used, the average sample number and operating characteristic of the Wald test with initial score $z$ satisfy the following Fredholm integral equations

of the second kind [11,31,32,44]:

$$\begin{aligned} \mathcal{M}(z,\theta) &= 1 + \int_0^h \mathcal{M}(x,\theta) dF(x-z|\theta) \\ \mathcal{P}(z,\theta) &= F(-z|\theta) + \int_0^h \mathcal{P}(x,\theta) dF(x-z|\theta) \end{aligned} \qquad z \in [0,h] \tag{8}$$

No analytical solutions can be found for these equations and we must resort to numerical techniques. We replace these integrals with an approximation according to

$$\begin{aligned} \widetilde{\mathcal{M}}(z,\theta) &= 1 + \sum_{j=1}^{J} w_j \widetilde{\mathcal{M}}(x_j,\theta) f(x_j - z|\theta) \\ \tilde{\mathcal{P}}(z,\theta) &= F(-z|\theta) + \sum_{j=1}^{J} w_j \tilde{\mathcal{P}}(x_j,\theta) f(x_j - z|\theta), \end{aligned} \tag{9}$$

where $\widetilde{\mathcal{M}}(z,\theta)$ and $\tilde{\mathcal{P}}(z,\theta)$ are the solutions, and the weights $\{w_j\}$ and points $\{x_j\}$ are chosen according to some rule. In [12], this approach is taken where the points $\{x_j\}$ are the "Gaussian points" or the roots of the Legendre polynomial and where the weights $\{w_j\}$ are the Gaussian coefficients [18]. Now, by evaluating (9) at the points $\{x_j\}$ which are substituted in for $z$, we obtain two systems of linear equations which can be solved by ordinary techniques to find the approximate solutions of the original integral equations. By substituting these values back in the right hand side of (9), $\widetilde{\mathcal{M}}(z,\theta)$ and $\tilde{\mathcal{P}}(z,\theta)$ can be evaluated for any values of $z$ of interest, and in particular for $z = 0$. This method of solution is called the Nyström approximation method for the solution of Fredholm integral equations of the second kind [5]. The matrix equations to be solved are of the form

$$(\mathbf{I} - \mathbf{A})\mathbf{u} = \mathbf{v} \tag{10}$$

where the matrix $(\mathbf{I} - \mathbf{A})$ has entries $(\mathbf{I} - \mathbf{A})_{ij} = \delta_{ij} - w_j f(x_i - x_j|\theta)$. Because of the introduction of the weights, this matrix has no special properties that allow efficient solution of the linear equations. Thus, they must be solved by a technique such as Gaussian elimination which requires that the number of multiplications for the solution is $1/3 J^3 + O(J^2)$ [24].

Instead of using the Gaussian points and weights, we can use a simple rectangular approximation to the integrals on the right side of (9). Again, this results in two sets of linear equations of the form (10) where $(\mathrm{I} - \mathrm{A})_{ij} = \delta_{ij} - \Delta f(x_i - x_j|\theta)$ for $\Delta = h/J$. This matrix is Toeplitz and there exists an efficient algorithm related to the celebrated Levinson's algorithm for the solution of the equations where the number of multiplications is $3J^2 + O(J)$ [27]. Moreover, the algorithm for the solution is a recursive one which computes the solution for lower order systems to arrive at the solution for the desired order. In other words, in solving for $\widetilde{\mathcal{P}}(0,\theta)$ and $\widetilde{\mathcal{M}}(0,\theta)$ for a given $h$, the solution is also computed for $h_i = i\Delta$ for $i = 1,\ldots,J-1$ with no extra computation. In contrast with this, the Nyström method requires that the linear equations must be solved anew for each different value of $h$ because the matrix is very general.

The results in [12] indicate that in certain situations only fifteen points may be required in the Nyström method for accuracy to six decimal places. It is well known that many more points are required for accurate solutions using the rectangular approximation. Our experimental results show that if very accurate solutions are required, the Nyström method is preferred. However, as pointed out in [12], for $1 - \mathcal{P}(0,\theta) < 10^{-5}$ the number of points may need to be increased. We have found that fifty to one hundred points may be needed to obtain a reasonable solution for large threshold values $h$ while the less accurate approach using the rectangular approximation may only require 250 points. Since the Nyström method must be repeated many times to specify the performance curve by calculating pairs of $(T, D)$ for many values of $h$, it may be worthwhile to consider the faster approach using the rectangular approximation. Although speed is not necessarily a factor in calculating the curves, they may take many hours of computation time for large values of $h$ for the *faster* method, so that one may want to consider speed over accuracy.

Both algorithms were implemented on a computer. It was found that the solution for $\tilde{\mathcal{P}}(0,\theta)$ became inaccurate as $h$ grew, to the extent that for $h$ large enough $\tilde{\mathcal{P}}(0,\theta)$ was computed to be larger than one. Since what is required to calculate $\mathcal{N}(0,\theta)$ is actually $\mathcal{Q}(0|\theta) = 1 - \mathcal{P}(0,\theta)$, the integral equation for $\mathcal{Q}(z,\theta)$ was utilized. Substituting $1 - \mathcal{Q}(z,\theta)$ into the integral equation for $\mathcal{P}(z,\theta)$ yields

$$\mathcal{Q}(z|\theta) = 1 - F(h - z|\theta) + \int_0^h \mathcal{Q}(x|\theta) dF(x - z|\theta).$$

$\mathcal{Q}(0|\theta)$ was computed according to both algorithms and the solutions were better behaved for larger values of $h$.

The procedure above could theoretically be applied when $g$ is an arbitrary nonlinearity by making the transformation of random variables, $Y = g(X)$, so that the resulting detector is again linear.

When the nonlinearity $g$ takes only the values in $\{-1, 0, +1\}$, the solution for $\mathcal{N}(0,\theta)$ is known exactly in closed form from results in random walk theory. Such a nonlinearity will be called a *random walk nonlinearity*. Examples of such nonlinearities are the *dead zone limiter* where

$$g(x) = \begin{cases} -1, & \text{for } x < -d \\ 0, & \text{for } -d \leq x < d \\ 1, & \text{for } x \geq d \end{cases}$$

and the sign detector which is obtained by setting $d = 0$. For the boundary $h$ a positive integer, the ASN of Page's test with a random walk nonlinearity is [26]

$$\mathcal{N}(0,\theta) = \begin{cases} \dfrac{h(1+h)}{p+q} & \text{if } p = q \\ \dfrac{1}{q-p}\left[\dfrac{1-(\frac{q}{p})^h}{\frac{p}{q}-1} - h\right] & \text{if } p \neq q \end{cases} \tag{11}$$

where $p = p(\theta)$ and $q = q(\theta)$ are defined by

$$p = P\{g(X) = 1|\theta\}$$

$$q = P\{g(X) = -1|\theta\}.$$

Thus, the performance curves are found by computing $T = \mathcal{N}(0, \theta_0)$ and $D = \mathcal{N}(0, \theta_1)$ for a range of boundaries $h = 1, 2, \ldots$.

To demonstrate the use of the performance curves, the following disorder problem is considered. The distributions before and after the disorder time are given by the symmetric signal in additive noise situation where $F(x|\theta_0) = F_0(x + \theta)$ and $F(x|\theta_1) = F_0(x - \theta)$ for $\theta > 0$. The noise distributions considered are the Gaussian with density

$$f_0(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/2\sigma^2),$$

the Laplace with density

$$f_0(x) = (2\sigma^2)^{-1/2} \exp(-\sqrt{2}|x|/\sigma),$$

and the Gaussian–Gaussian mixture with density

$$f_0(x) = (1 - \epsilon)\frac{1}{\sqrt{2\pi\sigma_0^2}}\epsilon^{-x^2/2\sigma_0^2} + \epsilon\frac{1}{\sqrt{2\pi\sigma_1^2}}\epsilon^{-x^2/2\sigma_1^2}.$$

The Gaussian–Gaussian mixture density is the first two terms in Middleton's Class A model [19,43] and has been used to model impulsive noise. The parameter $\epsilon$ indicates the degree of contamination and is typically in the range $(0, 0.25)$. The relative strength of the contamination is given by the power ratio $\gamma^2 \triangleq \sigma_1^2/\sigma_0^2$ which generally has values between 100 and 10,000. We use this distribution because by adjusting the parameters we can determine the performance for a wide range of distributions with different properties such as tail heaviness.

The nonlinearities implemented in Page's test are the linear detector $g(x) = x$, the sign detector, and the dead zone limiter. They are shown in Figure 2 along with optimal nonlinearities, the log-likelihood ratios, for signal-to-noise ratios (SNR) of 0 dB and $-20$ dB. The

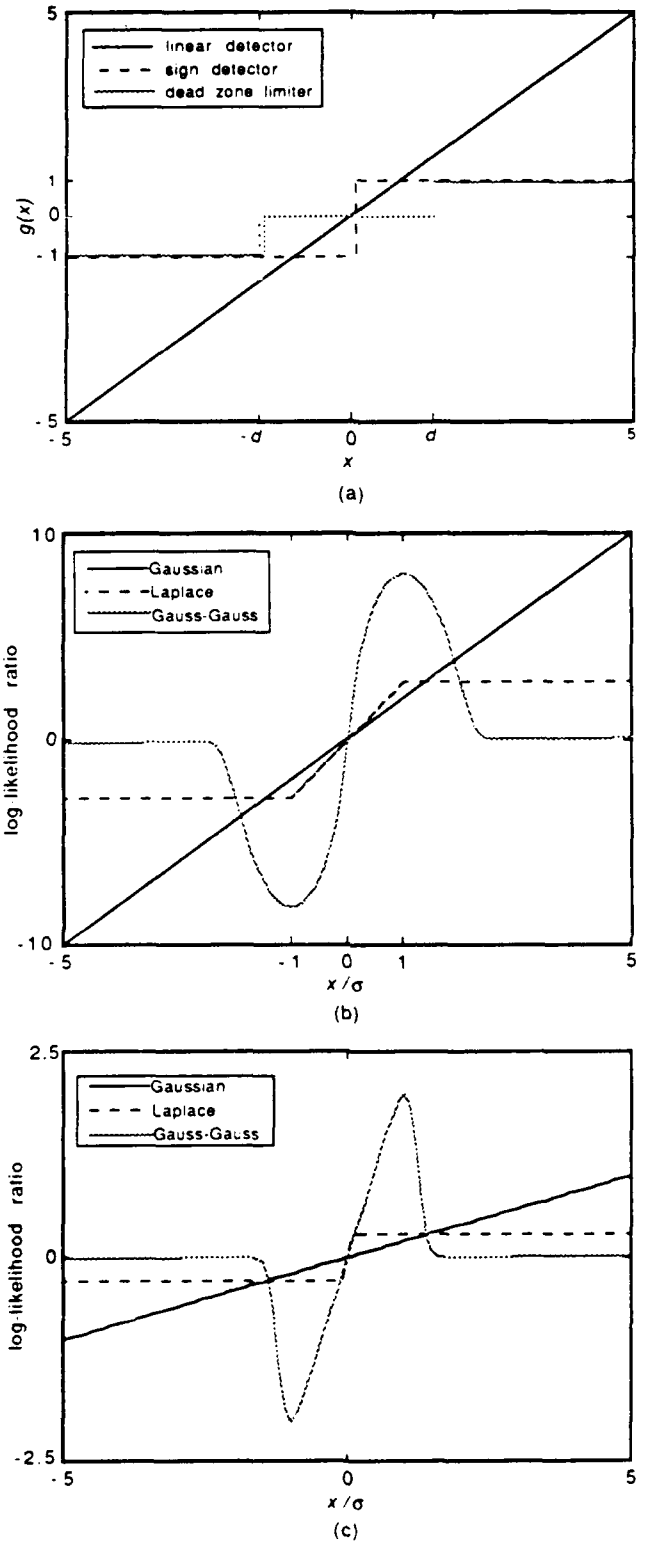



**Figure 2.** (a) The three nonlinearities used in Page's test and the log-likelihood ratios for Gaussian, Laplace, and the Gauss-Gauss mixture noise ($\epsilon = 0.01, \gamma^2 = 1000$) for an SNR of (b) 0 dB and (c) $-20$ dB.

dead zone limiter has the parameter $d$, which can be optimized for the particular distribution. However, since we assume that the true distribution is unknown, this optimization can not be done *a priori.* Instead, we use the optimal $d$ which maximizes the classical detector efficacy in the ordinary binary hypothesis testing problem for the Gaussian distribution regardless of the true distribution so that $d = 0.612\sigma$ [20]. Thus, we assume that the noise variance is known or can be estimated.

Figure 3 shows the performance curves of the three detectors for signal-to-noise ratios of 0 dB and $-20$ dB. The results for the linear detector were obtained using the rectangular approximation method. For the relatively large SNR of 0 dB, the performance is good in nearly all situations as shown in 3(a) — only a delay of a few samples for a mean time between false alarms on the order of 10,000 samples. The sign detector and dead zone limiter perform well for heavy-tailed distributions (the Laplace and the Gauss-Gauss mixture) whereas the linear detector can perform poorly. In Gaussian noise, the linear detector outperforms the others. These results are analogous with those obtained in the classical detection problem [20,42]. When the SNR is $-20$ dB, we see from Figure 3(b) that the performance is considerably degraded. Now, for a mean time between false alarms of 10,000 samples, the delay is on the order 100 to 200 samples. The performance curves can be explained qualitatively by comparing the nonlinearities implemented with the optimal nonlinearity, the log-likelihood ratio. The linear detector is optimal for Gaussian noise, to within an inconsequential scale factor. However, for the Laplace and mixture distributions, an important feature of the optimal nonlinearities is the limiting of outliers, a feature also present in the sign detector and the dead zone limiter.

There is an important observation we can make about the curves, independent of the detector and the underlying noise distribution. A common feature of all the performance

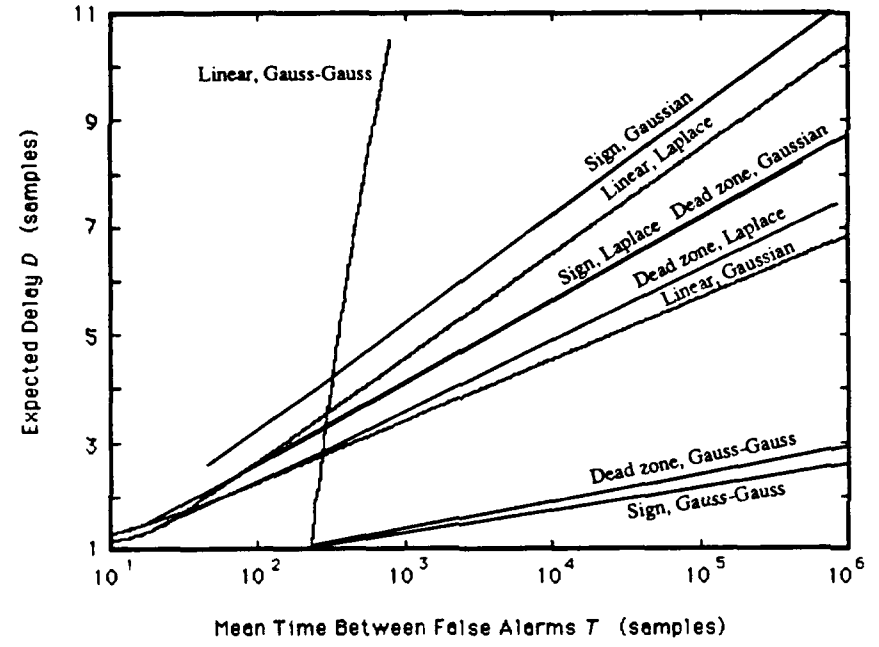



**Figure 3.** The performance curves for Page's test implemented with the linear detector, the sign detector, and the dead zone limiter operating in Gaussian, Laplace, and Gauss-Gauss mixture noise ($\epsilon = 0.01, \gamma^2 = 1000$) for a symmetric additive signal in noise with an SNR of (a) 0 dB and (b) $-20$ dB.

curves is that for large values of the mean time between false alarms $T$, the delay $D$ appears to be a linear function of the logarithm of $T$. Moreover, for large $T$, the ratio $\log(T)/D$ — which is the reciprocal of the slope — would be an appropriate performance measure. In the following section we clarify these statements.

## 2.3 Asymptotic Performance Measure

As suggested by the performance curves in the previous section we define an asymptotic performance measure $\eta$ for Page's test given by

$$\begin{aligned} \eta &\stackrel{\Delta}{=} \lim_{T\to\infty} \frac{\log T}{D} \\ &= \lim_{D\to\infty} \frac{\log T}{D} \\ &= \lim_{h\to\infty} \frac{\log T}{D} \end{aligned} \tag{12}$$

where the *natural* logarithm is used. The measure $\eta$ is the reciprocal of the asymptotic slope of the performance curve, and larger values of $\eta$ indicate better performance. In the remainder of this section, we will derive a lower bound for this measure and show that it has some interesting and useful properties.

At this point we review a result due to Lorden [25]. Let $L$ be the stopping variable of a one–sided test, that is,

$$L = \inf\{l : Z_l > h\},$$

for some statistics $\{Z_l\}$ which are functions of the i.i.d. observations $X_1, X_2, \ldots$. For $k = 1, 2, \ldots$, let $L_k$ be the stopping variable of the same test applied to $X_k, X_{k+1}, \ldots$ and define

$N$ as

$$N = \min_{k \geq 1}\{L_k + k - 1\}.$$

$N$ is the stopping variable of the test which stops when the first of the tests applied to $X_k, X_{k+1}, \ldots$ for $k = 1, 2, \ldots$ stops.

**Theorem.** (Lorden [25]) *Let $L$ be such that $P\{L < \infty|\theta_0\} \leq \alpha$. Then,*

$$E[N|\theta_0] \geq 1/\alpha,$$

*and for any alternative distribution,*

$$E[N|\theta_1] \leq E[L|\theta_1].$$

This theorem is relevant because Page's test with zero initial score is exactly the first Wald one–sided sequential test operating on $X_k, X_{k+1}, \ldots$ for $k = 1, 2, \ldots$ to stop as is shown in the example given in Figure 1(b). This can be seen from the view that when the test statistic $\tilde{S}_k$ would have fallen below zero — indicating that the test starting from $k+1$ will reach the boundary first — $\tilde{S}_k$ is reset to zero indicating that a new test is starting from $k + 1$.

To use the theorem, we examine the Wald one-sided test with stopping variable $L$ defined by

$$L = \inf\{m : S_m > h\}, \tag{13}$$

where

$$S_m = \sum_{i=1}^{m} g(X_i). \tag{14}$$

The probability that this test terminates, that is $P\{L < \infty\}$, is the probability that the boundary $h$ is exceeded. Although this can not be computed exactly for arbitrary jumps $g(X_i)$, an estimate is provided by the so–called *Wald approximations* and bounds [11,44].

First, we recall some results on the *non-zero* root $\omega = \omega(\theta)$ of the moment generating function identity

$$E[\exp\{\omega g(X)\}|\theta] = 1.$$

The existence and uniqueness of such a root is guaranteed when $g(X)$ has a non-zero mean and satisfies certain other conditions [11,44]. For our purposes, we will assume that a unique root exists. A key property of this root is that if $E[g(X)|\theta] > 0$ then the root is negative whereas if $g(X)$ has a negative mean, the root is positive [10]. When we are trying to detect a change from a negative trend to a positive one, we assume that $E[g(X)|\theta_0] < 0 < E[g(X)|\theta_1]$, implying that $\omega(\theta_1) < 0 < \omega(\theta_0)$.

In order to compute the probabilities and expectations required in the theorem for the one–sided test (13), we consider the two–sided Wald test with boundaries $b < 0 < h$ given by the stopping variable $M$,

$$M = \inf\{m : S_m < b \text{ or } S_m > h\}, \tag{15}$$

with $S_m$ as in (14). The one–sided test is now the limit of two–sided tests as $b$ tends to negative infinity, i.e.,

$$\lim_{b \to -\infty} M = L$$

in probability. Define the operating characteristic $\mathcal{P}(\theta)$ of the two–sided test according to

$$\mathcal{P}(\theta) = P\{S_M < b\}.$$

The following bounds for the OC will be used when $\omega(\theta) > 0$ [11,44]:

$$\frac{\exp\{h\omega(\theta)\} - 1}{\exp\{h\omega(\theta)\} - \kappa(\theta)\exp\{b\omega(\theta)\}} \leq \mathcal{P}(\theta) \leq \frac{\delta(\theta)\exp\{h\omega(\theta)\} - 1}{\delta(\theta)\exp\{h\omega(\theta)\} - \exp\{b\omega(\theta)\}} \tag{16}$$

where

$$\kappa(\theta) \triangleq \inf_{1<\xi<\infty} \xi E\left[\exp\{\omega(\theta)g(X)\} | \exp\{\omega(\theta)g(X)\} \leq \frac{1}{\xi}; \theta\right]$$

$$\delta(\theta) \triangleq \sup_{0<\xi<1} \xi E\left[\exp\{\omega(\theta)g(X)\} | \exp\{\omega(\theta)g(X)\} \geq \frac{1}{\xi}; \theta\right].$$

It follows that $\kappa(\theta) \leq 1$ and $\delta(\theta) \geq 1$. Also, we will use the following bounds for the average sample number $\mathcal{M}(\theta) = E[M|\theta]$ of the two-sided test [11,44]:

$$\mathcal{M}(\theta) \begin{cases} \geq \dfrac{b\mathcal{P}(\theta) + [h + \gamma(\theta)][1 - \mathcal{P}(\theta)]}{E[g(X)|\theta]} & \text{if E } [g(X)|\theta] < 0 \\ \geq \dfrac{b^2\mathcal{P}(\theta) + h^2[1 - \mathcal{P}(\theta)]}{E[g^2(X)|\theta]} & \text{if E } [g(X)|\theta] = 0 \\ \leq \dfrac{b\mathcal{P}(\theta) + [h + \gamma(\theta)][1 - \mathcal{P}(\theta)]}{E[g(X)|\theta]} & \text{if E } [g(X)|\theta] > 0 \end{cases} \tag{17}$$

where

$$\gamma(\theta) \triangleq \sup_{r>0} E[g(X) - r | g(X) \geq r; \theta].$$

The probability that the one-sided test (13) does not stop under $\theta_0$ is given by the limit as $b$ tends to negative infinity of the probability that the two-sided test (15) terminates at the lower boundary $b$:

$$\begin{aligned} P\{L = \infty|\theta_0\} &= \lim_{b\to-\infty} \mathcal{P}(\theta_0) \\ &\geq \lim_{b\to-\infty} \frac{\exp\{h\omega(\theta_0)\} - 1}{\exp\{h\omega(\theta_0)\} - \kappa(\theta_0)\exp\{b\omega(\theta_0)\}} \\ &= \frac{\exp\{h\omega(\theta_0)\} - 1}{\exp\{h\omega(\theta_0)\}}. \end{aligned}$$

Therefore, the probability that the one-sided test terminates is upper bounded by

$$P\{L < \infty|\theta_0\} \leq \exp\{-h\omega(\theta_0)\}.$$

An upper bound for the average sample number of the one-sided test $\mathcal{L}(\theta_1)$ can be obtained by using the upper bound for the ASN of the two-sided test $\mathcal{M}(\theta_1)$ and taking

the limit as $b$ tends to negative infinity:

$$\mathcal{L}(\theta_1) = E[L|\theta_1] = \lim_{b \to -\infty} \mathcal{M}(\theta_1)$$

$$\leq \lim_{b \to -\infty} \frac{b\mathcal{P}(\theta_1) + [h + \gamma(\theta_1)][1 - \mathcal{P}(\theta_1)]}{E[g(X)|\theta_1]}.$$

The right hand side is a decreasing function of $\mathcal{P}(\theta_1)$ so that the inequality is preserved if we replace $\mathcal{P}(\theta_1)$ by zero resulting in

$$\mathcal{L}(\theta_1) = E[L|\theta_1] \leq \frac{[h + \gamma(\theta_1)]}{E[g(X)|\theta_1]}.$$

Now, Lorden's theorem can be applied to yield results on Page's test with threshold $h$, increments $g(X)$, and zero initial score. The mean time between false alarms $T$ is lower bounded by

$$T = E[N|S_0 = 0, \theta_0] \geq \exp\{h\omega(\theta_0)\} \tag{18}$$

while the delay $D$ is upper bounded by

$$D = E[N|S_0 = 0, \theta_1] \leq \frac{[h + \gamma(\theta_1)]}{E[g(X)|\theta_1]}. \tag{19}$$

Using these results, we can lower bound the performance measure

$$\eta = \lim_{h \to \infty} \frac{\log T}{D}$$

$$\geq \lim_{h \to \infty} \frac{\log[\exp\{h\omega(\theta_0)\}]}{[h + \gamma(\theta_1)]/E[g(X)|\theta_1]}$$

$$= \omega(\theta_0)E[g(X)|\theta_1]$$

assuming that $\gamma(\theta_1)$ is finite. (This is a reasonable assumption and is obviously satisfied when $g$ is a bounded function.)

The same result can be obtained in a more direct manner by adopting the approach in [30]. Using equations (5) and (6), we see that the ASN of Page's test $\mathcal{N}(0, \theta)$ is the ratio

of the average sample number to the operating characteristic of a two–sided Wald test with boundaries 0 and $h$. Now, by using the bounds in (16) and (17), we can find bounds on $\mathcal{N}(0,\theta)$ directly by forming the desired ratio and taking the limit as $b$ tends to zero. We find a lower bound for the mean time between false alarms $T$ when $E[g(X)|\theta_0] < 0$:

$$\begin{aligned} T = \mathcal{N}(0,\theta_0) &= \lim_{b\uparrow 0} \frac{\mathcal{M}(\theta_0)}{1-\mathcal{P}(\theta_0)} \\ &\geq \lim_{b\uparrow 0} \frac{b\mathcal{P}(\theta_0) + [h+\gamma(\theta_0)][1-\mathcal{P}(\theta_0)]}{E[g(X)|\theta_0][1-\mathcal{P}(\theta_0)]} \\ &= \lim_{b\uparrow 0} \frac{1}{E[g(X)|\theta_0]} \frac{b\mathcal{P}(\theta_0)}{1-\mathcal{P}(\theta_0)} + \frac{h+\gamma(\theta_0)}{E[g(X)|\theta_0]}. \end{aligned}$$

The right hand side is a decreasing function of $\mathcal{P}(\theta_0)$ since $b < 0$. Thus, if we replace $\mathcal{P}(\theta_0)$ by its upper bound in (16) we get

$$T \geq \lim_{b\uparrow 0} \frac{b}{E[g(X)|\theta_0]} \frac{\delta(\theta_0)\exp\{h\omega(\theta_0)\} - 1}{1-\exp\{b\omega(\theta_0)\}} + \frac{h+\gamma(\theta_0)}{E[g(X)|\theta_0]}.$$

Both the numerator and denominator above are zero when $b = 0$, so we use L'Hopital's rule to find

$$\begin{aligned} T &\geq \frac{1}{E[g(X)|\theta_0]} \frac{1-\delta(\theta_0)\exp\{h\omega(\theta_0)\}}{\omega(\theta_0)} + \frac{h+\gamma(\theta_0)}{E[g(X)|\theta_0]} \\ &\geq \frac{1}{E[g(X)|\theta_0]} \frac{1-\exp\{h\omega(\theta_0)\}}{\omega(\theta_0)} + \frac{h+\gamma(\theta_0)}{E[g(X)|\theta_0]} \end{aligned}$$

where the last step follows from the fact that $\delta(\theta_0) \geq 1$. If $E[g(X)|\theta_0] = 0$, then we use (17) to obtain

$$\begin{aligned} T = \mathcal{N}(0,\theta_0) &= \lim_{b\uparrow 0} \frac{\mathcal{M}(\theta_0)}{1-\mathcal{P}(\theta_0)} \\ &\geq \lim_{b\uparrow 0} \frac{b^2\mathcal{P}(\theta_0) + h^2[1-\mathcal{P}(\theta_0)]}{E[g^2(X)|\theta_0][1-\mathcal{P}(\theta_0)]} \\ &= \lim_{b\uparrow 0} \frac{b^2\mathcal{P}(\theta_0)}{E[g^2(X)|\theta_0][1-\mathcal{P}(\theta_0)]} + \frac{h^2}{E[g^2(X)|\theta_0]}. \end{aligned}$$

The right hand side is an increasing function of $\mathcal{P}(\theta_0)$ so that if it is replaced by zero the inequality is preserved:

$$T \geq \frac{h^2}{E[g^2(X)|\theta_0]}.$$

The importance of this is that if $E[g(X)|\theta_0] = 0$, a significant penalty is paid. The mean time between false alarms is now a quadratic function of the threshold instead of an exponential one.

An upper bound for the expected delay $D$ is computed in the same manner as before using (17) where now $E[g(X)|\theta_1] > 0$:

$$\begin{aligned} D = \mathcal{N}(0,\theta_1) &= \lim_{b\uparrow 0} \frac{\mathcal{M}(\theta_1)}{1-\mathcal{P}(\theta_1)} \\ &\leq \lim_{b\uparrow 0} \frac{b\mathcal{P}(\theta_1) + [h+\gamma(\theta_1)][1-\mathcal{P}(\theta_1)]}{E[g(X)|\theta_1][1-\mathcal{P}(\theta_1)]} \\ &= \lim_{b\uparrow 0} \frac{1}{E[g(X)|\theta_1]} \frac{b\mathcal{P}(\theta_1)}{1-\mathcal{P}(\theta_1)} + \frac{h+\gamma(\theta_1)}{E[g(X)|\theta_1]}. \end{aligned}$$

The right hand side is a decreasing function of $\mathcal{P}(\theta_1)$ since $b < 0$. Thus, if we replace $\mathcal{P}(\theta_0)$ by zero the inequality is preserved and we obtain

$$D \leq \frac{h+\gamma(\theta_1)}{E[g(X)|\theta_1]}.$$

When $\omega(\theta_0) > 0$ and $\omega(\theta_1) < 0$, we have that as $h \to \infty$

$$\begin{aligned} T = \mathcal{N}(0,\theta_0) &\geq \frac{-e^{h\omega(\theta_0)}}{E[g(x)|\theta_0]\omega(\theta_0)} \\ D = \mathcal{N}(0,\theta_1) &\leq \frac{h}{E[g(x)|\theta_1]} \end{aligned}$$

so that the performance measure is lower bounded by

$$\eta = \lim_{h\to\infty} \frac{\log(T)}{D} \geq \omega(\theta_0) E[g(x)|\theta_1]. \tag{20}$$

This lower bound can also be shown to be an approximation to the actual value of $\eta$ by repeating the derivations above using the Wald approximations [11,44] instead of the bounds in (16) and (17).

The bound in (20) was derived for $E[g(X)|\theta_0] < 0$. However, if $E[g(X)|\theta_0] = 0$, we find that the performance measure $\eta$ is lower bounded by zero. This demonstrates that is

worthwhile to bias the detector when it is known that before the disorder occurs the detector output will have zero mean. If the detector is biased to have a negative mean before the disorder, then the mean time between false alarms will be an exponential function of the threshold instead of only a quadratic one. Thus, for a fixed mean time between false alarms, the delay for the biased detector will be on the order of the logarithm of the delay of the unbiased one. It should be noted, however, that the biased detector will not detect changes in mean, which, before the biasing, are less than the bias to be applied so that care must be taken in biasing the test.

Using the results in [21], the bound on the performance measure can easily be extended to apply to the so called "two-sided Page's test" in which the change in distribution is either from $F(\theta_0)$ to $F(\theta_1)$ or from $F(\theta_1)$ to $F(\theta_0)$. For simplicity, we consider only the ordinary version of Page's test with $E[g(X)|\theta_0] < 0 < E[g(X)|\theta_1]$.

## 2.4 Properties and Examples of the Performance Measure

Define the approximate asymptotic performance measure $\tilde{\eta}$ by

$$\tilde{\eta} = \omega(\theta_0)E[g(x)|\theta_1] \tag{21}$$

which is also a lower bound to the true value. $\tilde{\eta}$ has the following desirable properties.

*Property 1.* $\tilde{\eta}$ is invariant to scale changes in $g(x)$. Let $\tilde{g}(x) = cg(x)$ with $c > 0$. Then $\tilde{\omega}(\theta_0) = \omega(\theta_0)/c$ and $E[\tilde{g}(x)|\theta_1] = cE[g(x)|\theta_1]$. This is desirable because a scale change can be absorbed into the threshold and will not affect the performance. ■

*Property 2.* If $g$ is the log-likelihood ratio, $g(x) = \log[dF(x|\theta_1)/dF(x|\theta_0)]$, then the bound is tight, i.e., $\tilde{\eta} = \eta$. Lorden [25] proved that for the log-likelihood ratio $\eta = J(\theta_1;\theta_0)$ where

$$\begin{aligned} J(\theta_1;\theta_0) &= E\left[\log\frac{dF(X|\theta_1)}{dF(X|\theta_0)}\;\middle|\;\theta_1\right] \\ &= \int_{-\infty}^{\infty}\log\frac{dF(x|\theta_1)}{dF(x|\theta_0)}dF(x|\theta_1) \end{aligned}$$

which is called the *discrimination* or *Kullback–Leibler divergence* in information theory. Since

$$\begin{aligned} E\left[\exp\left\{\log\frac{dF(X|\theta_1)}{dF(X|\theta_0)}\right\}\;\middle|\;\theta_0\right] &= \int_{-\infty}^{\infty}\frac{dF(x|\theta_1)}{dF(x|\theta_0)}dF(x|\theta_0) \\ &= 1, \end{aligned}$$

it follows that $\omega(\theta_0) \equiv 1$. So,

$$\begin{aligned} \tilde{\eta} &= E[g(x)|\theta_1] \\ &= E\left[\log\frac{dF(X|\theta_1)}{dF(X|\theta_0)}\;\middle|\;\theta_1\right] \\ &= \eta. \end{aligned} \tag{22}$$

■

*Property 3.* When $g$ is a random walk nonlinearity, $\tilde{\eta} = \eta$: We can find $\eta$ explicitly by using the closed form solution for the average sample number in (10). Under the assumption that $E[g(x)|\theta_0] < 0$ and $E[g(X)|\theta_1] > 0$, we have that $q(\theta_0) > p(\theta_0)$ while $p(\theta_1) > q(\theta_1)$. Thus, the mean time between false alarms and the delay are asymptotically

$$\begin{aligned} T = \mathcal{N}(0,\theta_0) &\sim \frac{q(\theta_0)\left[\frac{q(\theta_0)}{p(\theta_0)}\right]^h}{[q(\theta_0)-p(\theta_0)]^2} \\ D = \mathcal{N}(0,\theta_1) &\sim \frac{h}{p(\theta_1)-q(\theta_1)} \end{aligned}$$

as $h \to \infty$. Thus, $\eta$ is given by

$$\eta = \lim_{h\to\infty}\frac{\log(T)}{D} = [p(\theta_1)-q(\theta_1)]\log\frac{q(\theta_0)}{p(\theta_0)}. \tag{23}$$

We have that

$$E\left[e^{\left(\log\frac{q(\theta_0)}{p(\theta_0)}\right)g(X)} \;\middle|\; \theta_0\right] = p(\theta_0)\frac{q(\theta_0)}{p(\theta_0)} + (1 - p(\theta_0) - q(\theta_0))\,e^0 + q(\theta_0)\frac{p(\theta_0)}{q(\theta_0)}$$
$$= 1$$

so

$$\omega(\theta_0) = \log\frac{q(\theta_0)}{p(\theta_0)}.$$

This, together with

$$E[g(X)|\theta_1] = p(\theta_1) - q(\theta_1),$$

yields the desired result. ■

*Property 4.* $\tilde{\eta} = \eta$ for the Brownian motion analog of the discrete-time Gaussian case using a continuous-time linear Page's test: Here, we observe the Brownian motion $B_t$ satisfying the stochastic differential equation

$$dB_t = \mu_t dt + \nu_t dw_t$$

where $w_t$ is a Wiener process, $\mu_t$ is the drift given by

$$\mu_t = \begin{cases} \theta_0, & \text{for } t < t_0 \\ \theta_1, & \text{for } t \geq t_0 \end{cases}$$

with $\theta_0 < 0 < \theta_1$, and $\nu_t$ is given by

$$\nu_t = \begin{cases} \sigma_0, & \text{for } t < t_0 \\ \sigma_1, & \text{for } t \geq t_0. \end{cases}$$

The continuous-time linear Page test is given by the stopping variable $U$ given by

$$U = \inf\{u : C_u > h\}$$

where $C_u$ is related to the observations by

$$C_u = B_u - \min_{0 \leq t \leq u} B_t.$$

The average sample number $\mathcal{U}(\theta,\sigma^2)$ of this test when $t_0 = \infty$ is given by [1,41]:

$$\mathcal{U}(\theta,\sigma^2) = E[U|\theta,\sigma^2)] = \begin{cases} \frac{h}{\sigma^2} & \text{if } \theta = 0 \\ \frac{1}{\theta}\left[\frac{\sigma^2}{2\theta}\left(e^{-2\theta h/\sigma^2} - 1\right) + h\right] & \text{if } \theta \neq 0. \end{cases}$$

Thus, the performance measure is

$$\eta = \lim_{h\to\infty} \frac{\log(T)}{D} = \frac{-2\theta_0\theta_1}{\sigma_0^2}.$$

This is the continuous-time analog of the discrete-time disorder problem with i.i.d. observations, $X_i \sim \mathrm{N}(\theta_0,\sigma_0^2)$ for $i < t_0$, $X_i \sim \mathrm{N}(\theta_1,\sigma_1^2)$ for $i \geq t_0$, with Page's test using $g(x) = x$. In this case, the moment generating function of $X$ before the disorder is

$$E[\exp\{tX\}|\theta_0,\sigma_0^2] = \exp\{\theta_0 t + \sigma_0^2 t^2/2\}$$

so that

$$\omega(\theta_0,\sigma_0^2) = -2\theta_0/\sigma_0^2.$$

This, and the fact that $E[X|\theta_1;\sigma_1^2] = \theta_1$ yields

$$\tilde{\eta} = \frac{-2\theta_0\theta_1}{\sigma_0^2} = \eta. \tag{24}$$

■

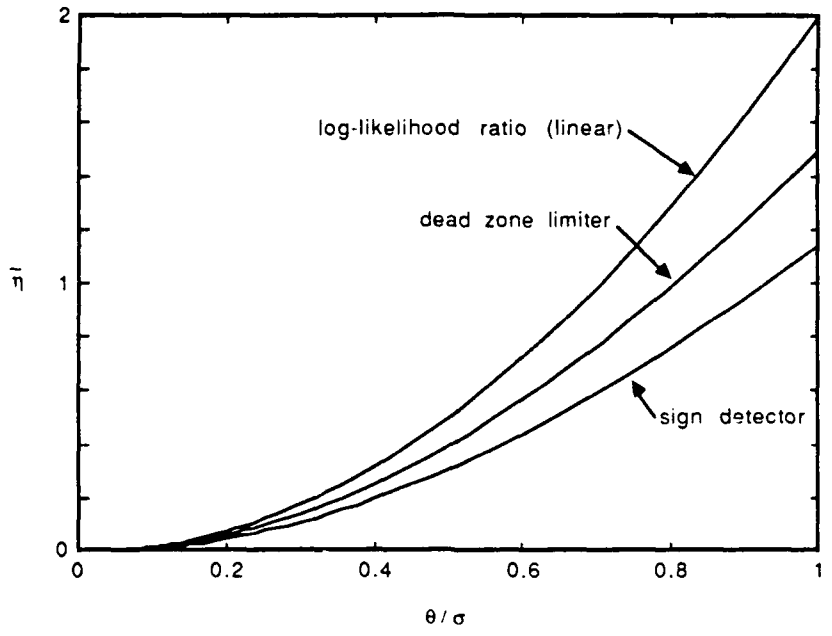
*Property 5.* In [16], Hinkley derived an asymptotic performance measure for the nonlinearity $g$ in an off-line formulation of the disorder problem using the probability of detection and probability of false alarm as the performance criteria. That measure $H$ is given by $H = \log E[\exp\{\omega(\theta_0)g(X)|\theta_1\}]$ whereas $\tilde{\eta}$ can be written as $\tilde{\eta} = E\left(\log[\exp\{\omega(\theta_0)g(X)|\theta_1\}]\right)$. By Jensen's inequality $H \geq \tilde{\eta}$. ■

Using $\tilde{\eta}$, the bound and approximate value of the performance measure, we can compare the performance of various nonlinearities in different noise distributions. Thus, we have an

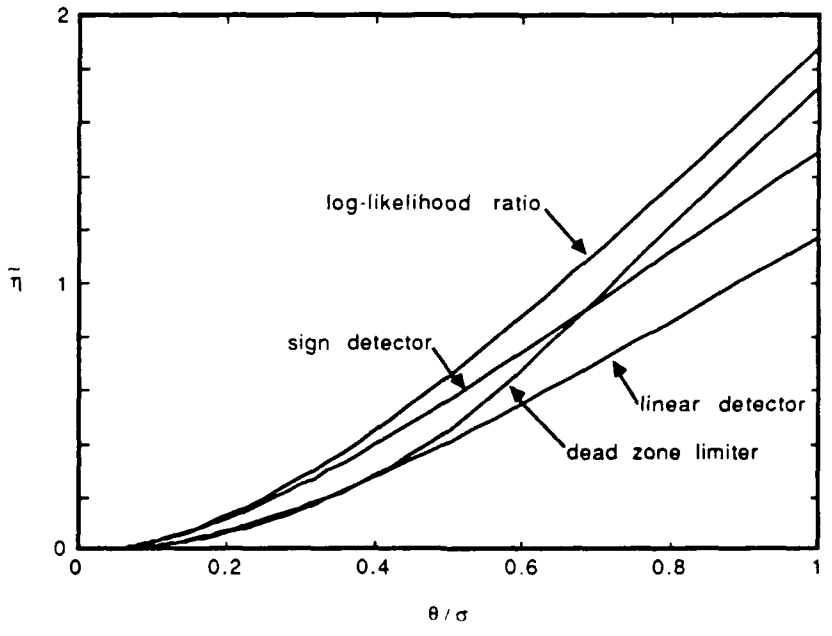
analytic measure and can avoid the complicated numerical integration discussed in Section 2.2 to generate the performance curves. Figure 4 shows the performance measure $\tilde{\eta}$ as a function of SNR for the linear detector, the sign detector, the dead-zone limiter ($d = 0.612\sigma$), and the log-likelihood ratio for Gaussian noise, Laplace noise, and Gauss-Gauss mixture noise with $\epsilon = 0.01$ and $\gamma^2 = 1000$. (Note: the results for the log-likelihood ratio for the Gauss-Gauss mixture could not be obtained numerically). In Figure 4(b) we see that the performance of the sign detector approaches that of the optimal detector for small SNR's. This is reminiscent of the result from classical detection theory that the sign detector is the locally optimal detector for Laplace noise. This will be examined in more detail in the next section. Figure 4(c) illustrates that the linear detector can perform very poorly for the Gauss-Gauss mixture where for an SNR of 0 dB the delay for the linear detector will be more than ten times larger than the delay for the sign detector.

The performance measure of the three detectors for the Gauss-Gauss mixture noise is shown as a function of $\epsilon$ for $\gamma^2 = 100$ and 1000 in Figure 5. We see that the sign detector and dead zone limiter can perform much better than the linear detector even though the linear detector is the optimal detector for ordinary Gaussian noise. We see from Figures 5(a) and 5(b) that the delay using the linear detector can be 100 times larger than the delay using the sign detector for certain mixture distributions.

The performance measure $\tilde{\eta}$ can be interpreted in two ways. We can use the ratio of two performance measures to indicate the relative delays for the two situations for a fixed mean time between false alarms large enough. For example, for small SNR's, Figure 4(b) indicates that using the linear detector in Laplace noise would result in twice the delay that the sign detector would have for a large mean time between false alarms. Alternatively, we can use $\tilde{\eta}$ to convert a large mean time between false alarms to the corresponding delay using the

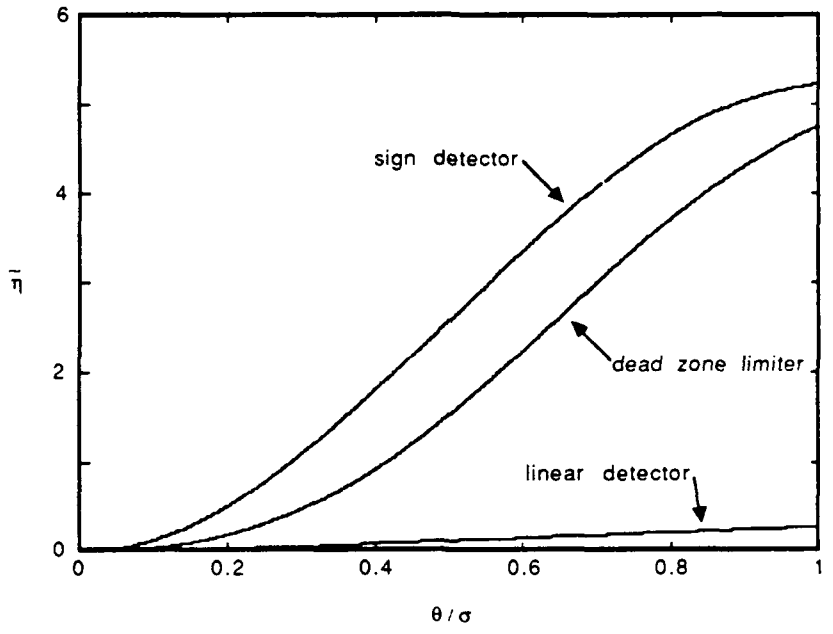



(a)

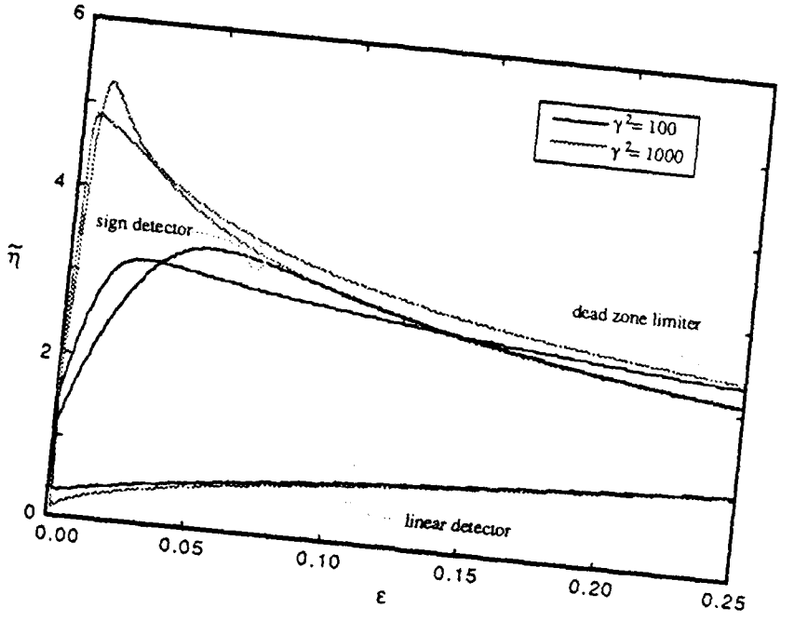



(b)

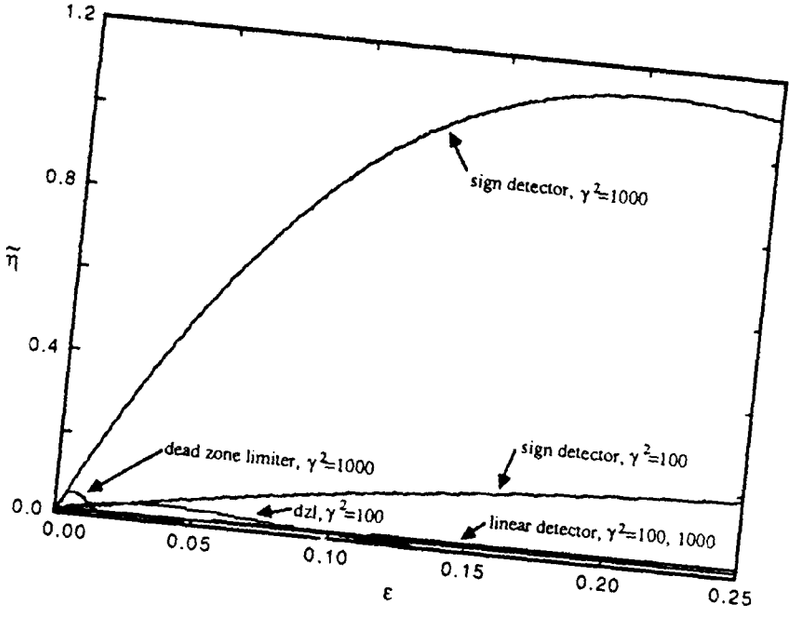
**Figure 4.** The performance measure $\tilde{\eta}$ as a function of the signal-to-noise ratio for Page's test implemented with the log-likelihood ratio, the linear detector, the sign detector, or the dead zone limiter operating in (a) Gaussian noise, (b) Laplace noise, and (c) Gauss-Gauss mixture noise with $\epsilon = 0.01$ and $\gamma^2 = 1000$.


6
4
2
0
0
0.2
0.4
0.6
0.8
1
$\tilde{\eta}$
sign detector
dead zone limiter
linear detector
θ / σ

(c)




(a)




(b)

**Figure 5.** The performance measure $\tilde{\eta}$ for Page's test implemented with the linear detector, the sign detector, and the dead zone limiter operating in Gauss-Gauss mixture noise with parameters $(\epsilon, \gamma^2)$ for a symmetric additive signal in noise with an SNR of (a) 0 dB and (b) −20 dB.

definition in (11). To demonstrate how one might use this interpretation of the performance measure, consider the following situation. Suppose we are observing the symmetric additive signal in noise disorder problem where the noise distribution is unknown but is assumed to be a Gauss-Gauss mixture with contamination parameter $\epsilon$ in the range $(0.05, 0.25)$ and $\gamma^2$ in the range $(100, 1000)$. The SNR is known to be 0 dB. Independent samples are taken at a rate of 10 kHz. It is desired to have less than one false alarm per hour, or less than one false alarm every $3.6 \times 10^7$ samples. We note from Figure 5(a) that $\tilde{\eta} \approx 3$ in the range of interest. The expected delay is upper bounded by $\log(T)/\tilde{\eta}$ which, for this example, is less than six samples or 600 microseconds. Similarly, for a false alarm rate of one per ten hours, the expected delay is less than seven samples or 700 microseconds.

## 2.5 Local Asymptotic Performance

The performance measure $\tilde{\eta}$ is a valid lower bound for all values of the parameters $\theta_0$ and $\theta_1$. Therefore, we investigate its properties in the so-called *local* or *weak signal* case. In addition to when the observations justify a weak signal treatment, the local performance is also important from a nonparametric viewpoint. Since we may not know the true signal value in a nonparametric setting, it is hoped that by using a good detector for the weak signal situation — which will not depend on the signal value — we can achieve good performance in the non-local situation as well.

Specifically, we consider the symmetric additive signal in noise situation, i.e., $F(x|\theta_0) = F_0(x + \theta)$ and $F(x|\theta_1) = F_0(x - \theta)$ where $F_0$ is some probability distribution and $\theta$ is

a positive parameter. Let $Y$ be a random variable with distribution $F_0$ which admits the probability density function $f_0(x)$. In addition to assuming that $E[g(Y-\theta)] < 0 < E[g(Y+\theta)]$ for all $\theta$ in a neighborhood of zero, it is assumed that $E[g(Y \pm \theta)]$ is continuous at $\theta = 0$ so that $E[g(Y)] = 0$. This implies that

$$\lim_{\theta \to 0} \tilde{\eta} = \lim_{\theta \to 0} \omega(-\theta) E[g(Y+\theta)] = 0.$$

This would be the case, for example, if $f_0$ is an even function and $g$ is an odd function. Thus, the local performance of Page's test is determined by the first non-zero derivative with respect to $\theta$ of $\tilde{\eta}$. Below, we derive the following key result under general smoothness and boundedness constraints:

$$\lim_{\theta \to 0} \frac{d}{d\theta} \tilde{\eta} = 0 \tag{25}$$

$$\lim_{\theta \to 0} \frac{d^2}{d\theta^2} \tilde{\eta} = 4 \frac{\{E[g'(Y)]\}^2}{E[g^2(Y)]} = 4\mathcal{E}. \tag{26}$$

Here, $\mathcal{E}$ is the efficacy of the detector $g$ in the classical detection problem. In the derivation, we will assume that limits and expectations can be interchanged freely, and similarly for derivatives and expectations. For a particular function $g$ and distribution $F_0$ these steps can be verified using the Lebesgue Dominated Convergence Theorem (see [16] for an example where $F_0$ is in an exponential family). We will further assume that all random variables considered are integrable, that is, the expectation exists and is finite. Again, whether this is true in a particular instance would need to be verified.

In deriving (25) and (26), we consider two cases: when $g(x)$ is the log-likelihood ratio and when $g(x)$ is a general nonlinearity which does not depend on $\theta$. In the former case, we have from (22)

$$\tilde{\eta} = E\{\_(Y+\theta)\} = E\left\{\log \frac{f_0(Y)}{f_0(Y+2\theta)}\right\}, \tag{27}$$

which is zero for $\theta = 0$. Under suitable regularity conditions, (27) can be differentiated with respect to $\theta$ yielding

$$\lim_{\theta \to 0} \frac{d}{d\theta} \tilde{\eta} = \lim_{\theta \to 0} E\left[\frac{-2f_0'(Y + 2\theta)}{f_0(Y + 2\theta)}\right] = -\int_{-\infty}^{\infty} 2f_0'(y)dy,$$

and this is zero because it is the derivative with respect to $s$ at $s = 0$ of the identity

$$\int_{-\infty}^{\infty} f_0(y + s)dy = 1. \tag{28}$$

Similarly, differentiating (27) twice with respect to $\theta$ and taking the limit yields

$$\lim_{\theta \to 0} \frac{d^2}{d\theta^2} \tilde{\eta} = -4 \int_{-\infty}^{\infty} f_0''(y)dy + 4 \int_{-\infty}^{\infty} f_0(y) \left[\frac{f_0'(y)}{f_0(y)}\right]^2 dy.$$

The first term can be seen to be zero by differentiating the identity (28) twice with respect to $s$ and setting $s = 0$, while the second term is four times the efficacy of the log-likelihood ratio, which coincides with Fisher's Information for the distribution $F_0$.

When $g(x)$ is a general nonlinearity which does not depend on $\theta$, it is assumed that $\omega(-\theta) \to 0$ as $\theta \to 0$. This is reasonable since the moment generating function $M(t) = E[\exp\{tg(Y - \theta)\}]$ is a convex function of $t$ with $M(t = 0) = 1$ and whose slope at $t = 0$ is equal to $E\{g(Y - \theta)\}$, which we assume increases to zero as $\theta$ increases to zero. Since $\omega(-\theta) \to 0$, the local performance of Page's detector is determined by the first non-zero derivative of $\tilde{\eta}$ with respect to $\theta$ at $\theta = 0$.

We have that $\tilde{\eta} = \omega(-\theta)E\{g(Y + \theta)\}$ which is differentiated to yield, under suitable regularity conditions,

$$\frac{d}{d\theta} \tilde{\eta} = E\{g(Y + \theta)\} \frac{d}{d\theta} \omega(-\theta) + \omega(-\theta)E[g'(Y + \theta)]. \tag{29}$$

Since $E[g(Y)] = 0$ and $\lim_{\theta \downarrow 0} \omega(-\theta) = 0$,

$$\lim_{\theta \downarrow 0} \frac{d}{d\theta} \tilde{\eta} = 0.$$

Differentiating (29) again yields

$$\frac{d^2}{d\theta^2}\eta = \frac{d^2}{d\theta^2}[\omega(-\theta)]E[g(Y+\theta)] + \frac{d}{d\theta}[\omega(-\theta)]E[g'(Y+\theta)]$$
$$+ \frac{d}{d\theta}[\omega(-\theta)]E[g'(Y+\theta)] + \omega(-\theta)E[g''(Y+\theta)]$$

Therefore, we have that

$$\lim_{\theta \downarrow 0} \frac{d^2}{d\theta^2}\tilde{\eta} = 2\lim_{\theta \downarrow 0} \frac{d}{d\theta}[\omega(-\theta)]E[g'(Y)]. \tag{30}$$

Next, the identity $1 = E[\exp\{\omega(-\theta)g(Y-\theta)\}]$ is differentiated twice with respect to $\theta$ and the limit is taken, resulting in

$$0 = E\{g^2(Y)\}\lim_{\theta \to 0}\left\{\frac{d}{d\theta}\omega(-\theta)\right\}^2 - 2E\{g'(Y)\}\lim_{\theta \to 0}\frac{d}{d\theta}\omega(-\theta).$$

This is solved yielding

$$\lim_{\theta \to 0}\frac{d}{d\theta}\omega(-\theta) = 2\frac{E\{g'(Y)\}}{E\{g^2(Y)\}} \quad \text{or} \quad \lim_{\theta \to 0}\frac{d}{d\theta}\omega(-\theta) = 0. \tag{31}$$

When the first relationship in (31) holds, (26) follows from (30).

The validity of these results is demonstrated in the following examples.

*Example 1.* When Page's test is implemented with the dead zone limiter, $\tilde{\eta}$ is given by (23)

$$\tilde{\eta} = \eta = [p(\theta_1) - q(\theta_1)]\log\frac{q(\theta_0)}{p(\theta_0)}.$$

We also assumed unbiasedness of the detector in the noise–only situation implying that $p(0) - q(0) = 0$. Since $p(\theta)$ and $q(\theta)$ are given by

$$p(\theta) = 1 - F_0(d-\theta)$$

$$q(\theta) = F_0(-d-\theta)$$

the unbiasedness condition implies

$$F_0(-d) = 1 - F_0(d)$$

where $[-d, d)$ is the *dead zone*. Computing $\frac{d}{d\theta}\tilde{\eta}$ yields

$$\frac{d}{d\theta}\tilde{\eta} = [f_0(d-\theta) + f_0(-d-\theta)]\log\left[\frac{F_0(-d+\theta)}{1-F_0(d+\theta)}\right] + [1 - F_0(d-\theta) - F_0(-d-\theta)]\left[\frac{f_0(-d+\theta)}{F_0(-d+\theta)} + \frac{f_0(d+\theta)}{1-F_0(d+\theta)}\right]$$

which is zero for $\theta = 0$ because of the unbiasedness condition. Computing the second derivative of $\tilde{\eta}$ results in

$$\begin{aligned}\frac{d^2}{d\theta^2}\tilde{\eta} = {} & [-f_0'(d-\theta) - f_0'(-d-\theta)]\log\left[\frac{F_0(-d+\theta)}{1-F_0(d+\theta)}\right] \\ & + [f_0(d-\theta) + f_0(-d-\theta)]\left[\frac{f_0(-d+\theta)}{F_0(-d+\theta)} + \frac{f_0(d+\theta)}{1-F_0(d+\theta)}\right] \\ & + [f_0(d-\theta) + f_0(-d-\theta)]\left[\frac{f_0(-d+\theta)}{F_0(-d+\theta)} + \frac{f_0(d+\theta)}{1-F_0(d+\theta)}\right] \\ & + [1 - F_0(d-\theta) - F_0(-d-\theta)]\frac{d}{d\theta}\left[\frac{f_0(-d+\theta)}{F_0(-d+\theta)} + \frac{f_0(d+\theta)}{1-F_0(d+\theta)}\right].\end{aligned}$$

For $\theta = 0$, this reduces to

$$\left.\frac{d^2}{d\theta^2}\tilde{\eta}\right|_{\theta=0} = \frac{2[f_0(d) + f_0(-d)]^2}{1 - F_0(d)}.$$

When the noise has a symmetric density, this becomes

$$\left.\frac{d^2}{d\theta^2}\tilde{\eta}\right|_{\theta=0} = \frac{8[f_0(d)]^2}{1 - F_0(d)}$$

which is four times the efficacy [20]. ■

*Example 2.* For Page's test implemented with the linearity $g(x) = x$, $\tilde{\eta}$ is given by (24), $\tilde{\eta} = 2\theta^2/\sigma^2$. The first derivative is $\frac{d}{d\theta}\tilde{\eta} = 4\theta/\sigma^2$ which is zero at $\theta = 0$. The second derivative is a constant, $\frac{d^2}{d\theta^2}\tilde{\eta} = 4/\sigma^2$, which is four times the efficacy of the linear detector in Gaussian noise [19]. ■

The fact that the performance measure is zero when there is no signal is expected and this implies that performance is poor for signal values near zero. The significance of the

fact that the first derivative of the performance measure is also zero at $\theta = 0$ is that the rate of improvement in performance is also poor for small signal levels. The importance of the second result relating the second derivative of the performance measure at $\theta = 0$ to the efficacy is that the wealth of results on the performance of detectors with cumulative sum test statistics in the classical hypotheses testing problem for detecting a weak additive signal in noise can be transfered to the weak signal performance of Page's test. For example, it is known that the so-called *locally optimal* nonlinearity $g(x) = -f_0'(x)/f_0(x)$ maximizes the efficacy in the classical detection problem. Therefore, it also maximizes the performance of Page's test in the small signal case.

## 2.6 Conclusion

In this chapter we have described a method to obtain the performance curves for Page's test with arbitrary nonlinearities. Perhaps more importantly, an asymptotic measure with a simple analytic formula was derived which can be useful in comparing different nonlinearities for a variety of noises in the desirable situation of a large mean time between false alarms. By using the Gaussian-Gaussian mixture family it was shown that the sign-based Page's test can outperform the linear Page's test for a wide range of distributions, which is a robustness property of value in nonparametric situations. Finally, by examining the local performance of Page's test, it was found that the performance is directly related to the efficacy of the classical binary hypothesis testing situation. Thus, the wealth of results on using memoryless nonlinearities in classical detection theory can be transferred to the quickest detection problem using Page's test.

## 2.7 References


[1] M. Basseville, "Edge detection using sequential methods for change in level – Part II: Sequential detection of change in mean," *IEEE Trans. Acous., Speech, Signal Processing*, vol. ASSP-29, no. 1, pp. 32–50, Feb. 1981.

[2] A. Benveniste, "Advanced methods of change detection: An overview," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, eds., Springer-Verlag, New York, pp. 77–99, 1986.

[3] G. K. Bhattacharyya and R. A. Johnson, "Nonparametric tests for shift at an unknown time point," *Ann. Math. Statist.*, vol. 39, no. 5, pp. 1731–1743, 1968.

[4] B. Broder and S. C. Schwartz, "The performance of Page's test and nonparametric quickest detection," *Proc. 1989 Conf. Information Sciences and Sytems*, Johns Hopkins University, Baltimore, MD, pp. 34–39.

[5] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, Van Nostrand Reinhold Co., New York, pp. 840–843, 1985.

[6] D. Brook and D. A. Evans, "An approach to the probability distribution of cusum run length," *Biometrika*, vol. 59, no. 3, pp. 539–549.

[7] H. Chernoff and S. Zacks, "Estimating the current mean of a normal distribution which is subjected to changes in time," *Ann. Math. Statist.*, vol. 35, pp. 999–1018, 1964.

[8] J. Deshayes and D. Picard, "Off-line statistical analysis of change-point models using non parametric and likelihood methods," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, eds., Springer-Verlag, New York, pp. 103–168, 1986.

[9] T. F. Dyson, *Topics in Nonlinear Filtering and Detection*, Ph.D. Dissertation, Dept. Elect. Eng., Princeton University, Princeton, N.J., pp. 73–107, 1986..

[10] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, New York, 1966.

[11] B. K. Ghosh, *Sequential Tests of Statistical Hypotheses*, Addison-Wesley, Reading, Mass., 1970.

[12] A. L. Goel and S. M. Wu, "Determination of A.R.L. and a contour nomogram for cusum charts to control normal mean," *Technometrics*, vol. 13, no. 2, pp. 221–230, May 1971.

[13] D. V. Hinkley, "Inference about the intersection in two-phase regression," *Biometrika*, vol. 56, no. 3, pp. 495–504, 1969.

[14] ——, "Inference about the change-point in a sequence of random variables," *Biometrika*, vol. 57, no. 1, pp. 1–17, 1970.

[15] ——, "Inference about the change-point from cumulative sum tests," *Biometrika*, vol. 58, no. 3, pp. 509–523, 1971.

[16] ——, "Time-ordered classification," *Biometrika*, vol. 59, no. 3, pp. 509–523, 1972.

[17] Z. Kander and S. Zacks, "Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points," *Ann. Math. Statist.*, vol. 37, pp. 1196–1210, 1966.

[18] L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis*, P. Noordhoff, Ltd., Groningen, The Netherlands, 1958.

[19] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, 1988.

[20] S. A. Kassam and J. B. Thomas, "Dead-zone limiter: An application of conditional tests in nonparametric detection," *J. Acoust. Soc. Am.*, vol. 60, no. 4, pp. 857–862, October 1976.

[21] R. A. Khan, "A note on Page's two-sided cumulative sum procedure," *Biometrika*, vol. 68, pp. 717–719.

[22] E. L. Lehmann, *Testing Statistical Hypotheses*, 2nd ed., Wiley, New York, 1986.

[23] ——, *Theory of Point Estimation*, Wiley, New York, 1983.

[24] J. S. Lim and A. V. Oppenheim, *Advanced Topics in Signal Processing*, Prentice Hall, Englewood, N.J., p. 18, 1988.

[25] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[26] C. A. Macken and H. M. Taylor, "On deviations from the maximum in a stochastic process," *SIAM J. Appl. Math.*, vol. 32, no. 1, pp. 96–104, 1977.

[27] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood, N.J., pp. 81–90, 1987.

[28] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Info. Theory*, vol. IT-18, pp. 241-250, March, 1972.

[29] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions." *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[30] I. V. Nikiforov, "Modification and analysis of the cumulative sum procedure," *Automation and Remote Control*, vol. 41, no. 9, part I, pp. 1247–1252, 1980.

[31] E. S. Page, "Continuous Inspection Schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.

[32] ——, "An improvement to Wald's approximation for some properties of sequential tests," *J. Royal Statist. Soc. B*, vol. 16, no. 1, pp. 136–139, 1954.

[33] ——, "A test for a change in a parameter occurring at an unknown point," *Biometrika*, vol. 42, pp. 523–526, 1955.

[34] A. N. Pettitt, "A nonparametric approach to the change point problem," *Appl. Statist.*, vol. 28, no. 2, pp. 126–135, 1979.

[35] ——, "A simple cumulative sum type statistics for the change-point problem with zero–one observations," *Biometrika*, vol. 67, no. 1, pp. 79–84, 1980.

[36] A. N. Shiryaev, "The detection of spontaneous effects," *Sov. Math. Dokl.*, no. 2, pp. 740–743, 1961.

[37] ——, "The problem of the most rapid detection of a disturbance in a stationary regime," *Sov. Math. Dokl.*, no. 2, pp. 795–799, 1961.

[38] ——, "On optimum methods in quickest detection problems," *Theory Prob. Appl.*, vol. 8, no. 1, pp. 22–46, 1963.

[39] ——, "Some exact formulas in a disorder problem," *Theory Prob. Appl.*, vol. 10, no. 3, pp. 348–354, 1965.

[40] A. F. M. Smith, "A Bayesian approach to inference about a change-point in a sequence of random variables," *Biometrika*, vol. 62, no. 2, pp. 407–416.

[41] H. M. Taylor, "A stopped Brownian motion formula," *Ann. Prob.*, vol. 3, no. 2, pp. 234–246, 1975.

[42] J. B. Thomas, "Nonparametric detection," *Proc. IEEE*, vol. 58, pp. 623–631, May 1970.

[43] K. S. Vastola, "Threshold detection in narrow-band non-Gaussian noise," *IEEE Trans. Comm.*, vol. 32, pp. 134–139, 1984.

[44] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.

# — 3 —

# The Application of Page's Test to Transient Signal Detection

## 3.1 Introduction

The problem of detecting unknown transient signals is an especially difficult one. The complications arise because of the presence of several unknown parameters which force the use of suboptimal detectors. These unknowns include the transient signal waveform and its amplitude, the duration of the transient, and the arrival time.

Several approaches have been taken for this problem. In [23], Wolcin derives a detector, robust for the transient signal waveform, and implements it sequentially to allow for arbitrary arrival times. Friedlander and Porat [7] use the Gabor representation to decompose a received waveform into a two-dimensional set of coefficients, one coordinate representing the complex frequency of a decaying sinusoid and the other representing the arrival time of the same signal. This approach is investigated in Chapter 4. Other work in transient signal detection and estimation can be found in [8,11,15,18], and an extensive bibliography can be

found in [12].

In this chapter, we consider a new approach which uses Page's test to detect transients. Although this test has been used in the related area of signal segmentation in conjunction with speech signals [2], the majority of the literature evaluates the test in the context of the disorder problem [3,5,17] as was done in Chapter 2. A disorder differs from a transient in that once the disorder occurs the change remains for all time, whereas a transient is usually thought of as having a finite duration. In the disorder problem, it has been shown that Page's test implementing the log-likelihood ratio nonlinearity is the quickest detector of the disorder [13.16]. Specifically, for a lower bounded mean time between false alarms, Page's test minimizes the worst case expected delay in detection once the disorder occurs.

The rationale for using Page's test to detect transient signals follows from the fact that it is the quickest detector for the disorder problem. Since the test detects the change as quickly as possible, once the detection occurs it is irrelevant that the disorder continues forever — it could just as well have been a transient that ended at the instant after the detection. Conversely, if another detector performed better for transients, one would suspect that it would be quicker than Page's test for the disorder problem, as well. However, one should be cautioned not to deduce that the optimality of Page's test for the disorder problem carries over to the transient problem. This can be understood by considering those realizations in the disorder problem which correspond to delays in detection larger than the duration of the transient. In such situations, the disorder is detected with some delay whereas the transient is not detected at all.

Another factor to consider in the question of the optimality of Page's test for transient signal detection is the problem of unknown parameters. Page's test is known to be the

quickest detector when the log-likelihood ratio is implemented in the test and the observations are independent and identically distributed (i.i.d.) with one distribution before the disorder and a second distribution after the disorder [13,16]. However, in the transient signal case, the i.i.d. assumptions may not hold, and even if they did, the distribution of the transient samples including all the parameters such as amplitude will not be known a priori. Thus, a suboptimal Page's test would have to be implemented, and other detection schemes might perform better. Despite these concerns about Page's test, the test does seem useful for detecting a change which occurs at an unknown time. Thus, the evaluation of the performance of Page's test for transient signal detection will be one of the major focuses of this chapter. In order to judge the performance of Page's test, several other tests found in the literature are also evaluated for the detection of transients. Previously, these tests had not been evaluated for transient signal detection in a sequential data processing framework.

The chapter is organized as follows. The use of Page's test for transient signal detection is justified in Section 2 by deriving a sequential algorithm to determine the maximum likelihood estimates of the starting and ending times of transients. Approximate methods of evaluating the performance of Page's test for transient signals are presented in Section 3. In Section 4 the performance is evaluated for the specific case of detecting a transient signal from observations of the sample energy spectral density. Alternatives to Page's test are presented in Section 5. Calculated and simulated results are discussed in Section 6 and conclusions are presented in Section 7.

## 3.2 Maximum Likelihood Estimation for Transients

In Chapter 2, Page's test was shown to be a sequential implementation of a maximum likelihood detector under the assumption that the change in the underlying process can be modeled as a disorder process. However, Page's test has only been considered when it is assumed that once the disorder occurs, the new probabilistic model remains for all time thereafter [eg. 1,5,17]. This is obviously not the case for transients, for although they occur abruptly at an unknown time, they also disappear (or fade away) at some later time. By deriving the maximum likelihood estimates for the starting and ending times of the transient, we will show that the resulting detector also implements Page's test.

We assume the simplest possible model for the transient signal in noise. Let the transient begin at time $t_0$ and end at time $t_1$. Before $t_0$ and after $t_1$, the noise is modeled by an i.i.d. random process with distribution $F_0$. During the transient, the signal *together* with the noise is modeled by an i.i.d. random sequence with distribution $F_1$. Thus, if $X_1, X_2, \ldots, X_N$ is the fixed set of observations, then the situation is as follows:

$$X_n \sim \text{i.i.d. } F_0, \qquad 1 \le n < t_0$$

$$X_n \sim \text{i.i.d. } F_1, \qquad t_0 \le n \le t_1$$

$$X_n \sim \text{i.i.d. } F_0, \qquad t_1 < n \le N$$

The maximum likelihood estimates of $t_0$ and $t_1$ are given by

$$\begin{aligned} t_0, t_1 &= \arg \max_{t_0, t_1} \left\{ \sum_{n=1}^{t_0-1} \log f_0(X_n) + \sum_{n=t_0}^{t_1} \log f_1(X_n) + \sum_{n=t_1+1}^{N} \log f_0(X_n) \right\} \\ &= \arg \max_{t_0, t_1} \left\{ \sum_{n=t_0}^{t_1} \log \frac{f_1(X_n)}{f_0(X_n)} + \sum_{n=1}^{N} \log f_0(X_n) \right\}. \end{aligned}$$

We now consider the maximization of the likelihood function as indicated above, that is

$$\max_{1 \le t_0 \le t_1 \le N} \left\{ \sum_{n=t_0}^{t_1} \log \frac{f_1(X_n)}{f_0(X_n)} \right\}, \tag{1}$$

ignoring the term which does not depend on the transient epochs. In the maximization of (1), we include the case that the sum contains no terms, which corresponds to the situation in which no transient is present. For this case, the sum is by definition equal to zero, and, therefore, the maximum in (1) is nonnegative. Thus, we compute

$$\max_{1\le t_0\le t_1\le N}\left\{\left(\sum_{n=t_0}^{t_1}\log\frac{f_1(X_n)}{f_0(X_n)}\right)^+\right\},$$

where $(\bullet)^+ = \max\{0,\bullet\}$. We derive a sequential algorithm for performing the maximization and show that it is, indeed, an implementation of Page's test. Assume the maximum has been computed for $N = n$, and let $\gamma_n$ be that maximum which is given by

$$\gamma_n = \max_{1\le t_1\le n}\left\{\max_{1\le t_0\le t_1}\left(\sum_{m=t_0}^{t_1}\log\frac{f_1(X_m)}{f_0(X_m)}\right)^+\right\}.$$

The maximum for $N = n+1$, $\gamma_{n+1}$, can be computed from

$$\begin{aligned}\gamma_{n+1} &= \max_{1\le t_1\le n \text{ or } t_1=n+1}\left\{\max_{1\le t_0\le t_1}\left(\sum_{m=t_0}^{t_1}\log\frac{f_1(X_m)}{f_0(X_m)}\right)^+\right\} \qquad (2)\\ &= \max\left\{\max_{1\le t_1\le n}\left[\max_{1\le t_0\le t_1}\left(\sum_{m=t_0}^{t_1}\log\frac{f_1(X_m)}{f_0(X_m)}\right)^+\right], \max_{1\le t_0\le n+1}\left(\sum_{m=t_0}^{n+1}\log\frac{f_1(X_n)}{f_0(X_n)}\right)^+\right\}.\end{aligned}$$

Let $\delta_n$ and $\lambda_n$ be given by

$$\begin{aligned}\delta_n &\triangleq \max_{1\le t_0\le n}\left(\sum_{m=t_0}^{n}\log\frac{f_1(X_m)}{f_0(X_m)}\right)^+\\ \lambda_n &\triangleq \log\frac{f_1(X_n)}{f_0(X_n)}\end{aligned}$$

so that

$$\delta_{n+1} = (\delta_n + \lambda_n)^+.$$

From (2) it is clear that

$$\gamma_{n+1} = \max\{\gamma_n, \delta_{n+1}\}$$

so that the sequential algorithm for computing the maximum of the likelihood function is:

$$\begin{aligned}
\delta_0 &= 0 \\
\gamma_0 &= 0 \\
\lambda_n &= \log \frac{f_1(X_n)}{f_0(X_n)} \\
\delta_n &= \max\{0, \delta_{n-1} + \lambda_n\} \qquad\qquad n = 1, 2, \ldots, N \\
\gamma_n &= \max\{\gamma_{n-1}, \delta_n\}.
\end{aligned}$$

The statistic $\gamma_n$ is an indication of the likelihood of a transient signal of some length in the interval $[1, n]$ — the larger the statistic, the more likely a transient is present. A reasonable detection scheme would then be

$$\begin{array}{lll} \gamma_n \overset{>}{<} h & \Rightarrow & \text{Declare Transient Signal Present} \\ & \Rightarrow & \text{Continue, } n = n + 1. \end{array} \tag{3}$$

Note that $\delta_n$ is identical to Page's statistic $\tilde{S}_n$ in (2.3), and that $\gamma_n$ can be rewritten

$$\gamma_n = \max\{\delta_0, \delta_1, \ldots, \delta_n\}.$$

This implies that the test (3) is equivalent to Page's test (2.4).

By not integrating over the observations which have a greater probability of containing only noise than noise plus signal, Page's test has desirable "windowing" properties which are suited for detection problems when events occur at unknown times. The effect of resetting Page's statistic to zero instead of allowing it to fall below zero is to forget all of the past observations. In such a situation the past is not relevant because it becomes more likely that a transient has not yet occurred. This resetting has the effect of windowing the data so that the noise-only observations before the transient occur are ignored. A detection will then occur if the signal duration and/or signal-to-noise ratio are large enough so that the test statistic exceeds the threshold.

## 3.3 Performance

We would like to determine the performance of the proposed test. To do so, we begin by evaluating the performance for the disorder problem where the quickest detection performance criteria are relevant. These criteria are the mean time between false alarms and the expected delay. Then, using these results, we relate performance in the disorder problem to performance in the transient problem. The mean time between false alarms is easily related to the false alarm rate, but the same can not be said for the expected delay and the probability of detection. Several approaches will be given to determine the probability of detection and relate it to the performance measures developed for quickest detection procedures.

For convenience, we parameterize the distributions according to $F_0 = F(\theta_0)$ and $F_1 = F(\theta_1)$. From Chapter 2, the mean time between false alarms $T$ can be lower bounded by the Lorden bound (2.18)

$$T \geq \exp\{\omega(\theta_0)h\} \tag{4}$$

where $\omega(\theta_0)$ is the unique non-zero root (which is assumed to exist) of the moment generating function identity

$$E[\exp\{\omega g(X)\}|\theta] = 1 \tag{5}$$

for the nonlinearity $g$. Also, using the Wald approximations we have that

$$T \approx \frac{1}{E[g(X)|\theta_0]}\left[\frac{1-\exp\{h\omega(\theta_0)\}}{\omega(\theta_0)} + h\right]. \tag{6}$$

The false alarm rate $\alpha$ is related to the mean time between false alarms by the straightforward relation

$$\alpha = \frac{1}{T} \tag{7}$$

which has units "false alarms per sample $X_i$" ( note: $X_i$ can represent one random variable or a vector of random variables). Because we use a one-sided sequential algorithm, we avoid the confusing terminology "the probability of false alarm" which is used for fixed sample size tests. Indeed, in a sequential test such as ours the probability of false alarm is exactly one because, if no disorder occurs, the test will terminate at the boundary $h$ if one waits long enough. The critical information, however, is the mean time between false alarms or equivalently the false alarm rate. The worst case expected delay can be upper bounded by (2.19)

$$D \leq \frac{[h + \gamma(\theta_1)]}{E[g(X)|\theta_1]}, \tag{8}$$

or, by ignoring the excess over the boundary, we have that

$$D \approx \frac{h}{E[g(X)|\theta_1]}. \tag{9}$$

In Chapter 2 it was shown that an appropriate performance measure for Page's test is the reciprocal of the asymptotic slope of the performance curve indicating $D$ versus $\log(T)$ as $T \to \infty$. This measure, $\eta$, is lower bounded and approximated by $\tilde{\eta}$ given by (2.21)

$$\tilde{\eta} = \omega(\theta_0)E[g(X)|\theta_1] \leq \eta = \lim_{T\to\infty} \frac{\log(T)}{D}. \tag{10}$$

The expected delay and the performance measure are valuable criteria for the disorder problem, but how can they be related to the probability of detection for the transient problem? Indeed, in the disorder problem the probability of detection is exactly one because it would eventually exceed the threshold if allowed to run long enough. This is obviously not meaningful in the transient situation where a change occurs only for a short time.

Some simple bounds on the probability of detection of the transient can be obtained by applying Markov's inequality on the positive integer-valued random variable $L$. As

mentioned above, the worst case expected delay can be calculated by

$$D = E[L|\theta_1, S_0 = 0].$$

Since

$$(n+1)I_{\{L>n\}} \leq L,$$

where $I_A$ is the indicator random variable of the measurable set $A$, it follows that

$$\mathcal{P}\{L > n|\theta_1, S_0 = 0\} \leq \frac{E[L|\theta_1, S_0 = 0]}{n+1} = \frac{D}{n+1}. \tag{11}$$

Let $\tau$ be the integer-valued duration of the transient and let $\beta$ be the probability of detecting the transient. A detection will be defined as a test which terminates *during the life of the transient*. Thus, a test which terminates on the sample after the transient is over will be considered a false alarm and not a detection. This is done to avoid ambiguities and the question of until how long after the transient has ceased should "detections" be attributed to the transient. The probability of detection is a function of the state of the cumulative sum statistic at the moment before the transient occurs so that

$$\beta = \beta(z) = \mathcal{P}\{L \leq \tau|\theta_1, S_0 = z \geq 0\}.$$

$\beta$ is lower bounded by the worst case scenario in which $S_0 = 0$ so that (11) can be used to find

$$\begin{aligned} \beta &\geq \mathcal{P}\{L \leq \tau|\theta_1, S_0 = 0\} \\ &\geq \frac{\tau + 1 - D}{\tau + 1}. \end{aligned} \tag{12}$$

This formula is useful if the expected delay is small in comparison with the transient duration, and it also illustrates the obvious fact that for the disorder problem, where $\tau = \infty$, the probability of detection is one when the expected delay $D$ is finite. Using (8) and (10), we have that

$$D \leq \left(\frac{|\log \alpha|}{\tilde{\eta}}\right)\left(\frac{h + \gamma(\theta_1)}{h}\right),$$

which can be substituted into (12) to lower bound the receiver operating characteristic, or, for large $D$,

$$\beta \geq \frac{\tau + 1 + \frac{\log \alpha}{\tilde{\eta}}}{\tau + 1}. \tag{13}$$

When the expected delay is large, asymptotic results on the distribution of the run length can be used to obtain an approximation to the probability of detection. Set $\mu = E[g(X)|\theta_1]$ and $\sigma^2 = \mathrm{Var}[g(X)|\theta_1]$. The asymptotic distribution of the run length under $F(\theta_1)$ is as follows [10]

$$\frac{(L - \frac{h}{\mu})}{\sigma\sqrt{h/\mu^3}} \xrightarrow{\mathcal{L}} N(0,1)$$

as $h \to \infty$, where $N(a,b)$ is the Gaussian distribution with mean $a$ and variance $b$. Using the asymptotic distribution, the probability of detection can be approximated by

$$\beta = \mathcal{P}\{L \leq \tau\} \approx \Phi\left(\frac{\tau - h/\mu}{\sigma\sqrt{h/\mu^3}}\right) \tag{14}$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-x^2/2} dx.$$

As done above, substitutions can be made for $h$ and $\mu$ with $\alpha$, $D$, and $\tilde{\eta}$ to obtain additional relationships from (14).

The most useful approximation can be obtained by using the well known result [6] that $\tilde{S}_n$ has the same distribution as $M_n$ where

$$M_n = \max\{0, S_1, S_2, \ldots, S_n\},$$

and where $\tilde{S}_n$ and $S_1, S_2, \ldots, S_n$ are related through (2.1) and (2.3). Again, we assume the worst case formulation and shift the time axis so that the transient occurs at time $n = 1$

and that $S_0 = \tilde{S}_0 = 0$. Then the actual probability of detection is lower bounded by

$$\begin{aligned}
\beta &\geq \mathcal{P}\left\{\tilde{S}_n > h \text{ for some } n = 1, 2, \ldots, \tau\right\} \\
&\geq \mathcal{P}\{\tilde{S}_\tau > h\} \\
&= \mathcal{P}\{M_\tau > h\} \\
&= \mathcal{P}\{S_n > h \text{ for some } n = 1, 2, \ldots, \tau\} \qquad (15) \\
&\geq \mathcal{P}\{S_\tau > h\}. \qquad (16)
\end{aligned}$$

The lower bound has been reduced several times, but the resulting lower bound is both simple and, as we will see later, quite useful. The threshold $h$ can be upper bounded by

$$h \leq \frac{|\log \alpha|}{\omega(\theta_0)},$$

so that the receiver operating characteristic is lower bounded by

$$\beta \geq \mathcal{P}\left\{S_\tau > \frac{|\log \alpha|}{\omega(\theta_0)}\right\}.$$

The probability (16) can be computed explicitly in certain cases. For example, when $X_1, X_2, \ldots$ are Gaussian random variables then so is $S_\tau$ and the probability is easily computed. Also, even when the distribution of the sum of the underlying random variables can not be found explicitly, the probability can be approximated for large $\tau$ using the Central Limit Theorem.

Another approximation can be obtained by modeling the cumulative sum statistic by a Brownian motion in continuous time. Let $W(t)$ be a Wiener process, $h(t) = a + bt$ for $a \geq 0$ and $b \leq 0$, and define a stopping variable $T(h)$ by

$$T(h) = \inf\{t > 0 | W(t) > h(t)\}.$$

The distribution of $T(h)$ is known to be [13]

$$\mathcal{P}\{T(h) < u\} = \int_0^u \frac{a}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(a+bt)^2}{2t}\right\} dt.$$

Using (15), we have that

$$\beta \geq \mathcal{P}\{\hat{T}(h) \leq \tau\},$$

where $\hat{T}(h)$ is defined by

$$\hat{T}(h) = \inf\{t > 0 | S_n > h\}.$$

Now, since $E[S_n] = n\mu$ and $\mathrm{Var}[S_n] = n\sigma^2$, we make the identifications:

$$\frac{S_n - n\mu}{\sigma} \longleftrightarrow W(t)$$
$$\frac{h - n\mu}{\sigma} \longleftrightarrow a + bt.$$

Thus, the probability of detection is approximately

$$\begin{aligned}\beta &\geq \mathcal{P}\{\hat{T} \leq \tau\} \\ &\approx \int_0^{\tau+\frac{1}{2}} \frac{h}{\sqrt{2\pi\sigma^2 t^3}} \exp\left\{-\frac{(h-\mu t)^2}{2t\sigma^2}\right\} dt.\end{aligned}$$

One would expect this approximation to be better when the i.i.d. summands forming $S_n$ are Gaussian random variables or when $\tau$ is large.

In this section, we have developed a number of bounds on the false alarm rate and probability of detection when Page's test is used for transient signal detection. Some of these bounds are derived from and related to the quickest detection performance criteria for the disorder problem. In particular, the simple bounds in (4) and (16) will be used in conjunction with the Monte Carlo simulations in Section 6.

## 3.4 Page's Test Applied to the Energy Spectral Density

We are interested in detecting transient signals using the sample energy spectral density (ESD) as the observations. In this regard, we are motivated by the work of Wolcin [23]. A detailed description of his method is given in the Appendix, and, in the simulations in Section 6, that method will be compared with the one we develop here.

When no transients are present, so that we are in the noise only situation, we assume that the background noise is *white Gaussian noise* (WGN). We assume that the energy spectral density is white merely for analytical convenience. For a colored spectrum, the approach of Wolcin [23] is to generate an unbiased estimate of the true spectrum and to use this to whiten (ie., normalize) the spectrum. The same approach can be used here so that we make the ideal assumption that the spectrum is white. The Gaussian observations are grouped in blocks of $N$ points for processing via the *Discrete Fourier Transform* (DFT). The squared magnitudes of the $N$ complex outputs of the DFT are computed, and these statistics are available for the detection procedure. These random variables will be denoted by $\{X_{in}\}$, $i = 1, 2, \ldots,$ $n = 1, 2, \ldots, N$ where $i$ is the block number and $n$ is the frequency bin number.

The nonparametric spectral estimator described above is called the Periodogram spectral estimator and is widely used because of its simplicity. In [1,22], a more complicated nonparametric spectral estimator based on the theory of *evolutionary spectra* due to Priestly [19,20] is used in conjunction with Page's test to detect an abrupt change in the spectrum of a piecewise stationary process. However, the performance of these tests were not evaluated in the context of transient detection. In [4], an autoregressive moving average model is assumed for the data but no evaluation of detection performance is given.

Under the white Gaussian noise assumption, the random variables $\{X_{in}\}$ will be independent and identically distributed with the exponential distribution and unity mean. (Note: this ignores the fact that the distribution of the first and last frequency bins will actually have the $\chi_1^2$ distribution with a variance twice that of the other bins [20], or it can be assumed that those two frequency bins are not used.) Thus, each sample ESD will be independent of the others and will have a probability density $p_n$ given by

$$p_n(x_{i1}, x_{i2}, \ldots, x_{iN}) = \prod_{n=1}^{N} \exp\{-x_{in}\} = \exp\left\{-\sum_{n=1}^{N} x_{in}\right\}$$

for $x_{in} \geq 0$.

The transient signal we are attempting to detect begins at the unknown block $t_0$ and continues through the block $t_1$. As in [23], we make the following ideal assumptions about the sample ESD obtained when the transient is present: 1) the random variables corresponding to different frequency bins are mutually independent, and 2) the sample ESD's corresponding to different time blocks are mutually independent, as well. The distribution of an ESD sample with a transient signal present (in addition to the WGN) will also be assumed to be exponential but now with a mean $\mu$ which is greater than one, $f_\mu(x) = \mu^{-1} \exp(-x/\mu)$ for $x \geq 0$. This would be the case, for example, when the transient signal is also a Gaussian random process which is independent of the background noise process. The probability density $p_s$ of the spectral samples for block $i$ for $t_0 \leq i \leq t_1$ can now be written

$$p_s(x_{i1}, x_{i2}, \ldots, x_{in}) = \prod_{n=1}^{N} \frac{1}{\mu_{in}} \exp\{-x_{in}/\mu_{in}\} = \exp\left\{-\sum_{n=1}^{N} x_{in}/\mu_{in} + \log \mu_{in}\right\} \quad (17)$$

for $x_{in} \geq 0$. Here, $\mu_{in} \geq 1$ with strict inequality if the transient signal is present in bin $(i, n)$. This model is as general as Wolcin's in allowing for different power levels $\{\mu_{in}\}$ in each time/frequency bin.

Because the parameters $\{\mu_{in}\}$ are not known a priori and because the distribution of

the $i$'th ESD need not be identical with the $j$'th ESD for $t_0 \leq i, j \leq t_1$, Page's test using the optimal nonlinearity, the log-likelihood ratio, can not be implemented. If, on the other hand, the $\{\mu_{in}\}$ were known *and* did not depend on $i$ so that $\mu_{in} = \mu_n$ for $t_0 \leq i \leq t_1$, then Page's test would be

$$\begin{array}{ll} \tilde{S}_i \begin{array}{l} > \\ < \end{array} h & \begin{array}{l} \Rightarrow \\ \Rightarrow \end{array} \begin{array}{l} \text{Declare Transient Signal Present} \\ \text{Continue, } i = i + 1 \end{array} \end{array}$$

where

$$\tilde{S}_i = \max\{0, \tilde{S}_{i-1} + g(X_{i1}, X_{i2}, \ldots, X_{iN})\}$$

$$\tilde{S}_0 = 0$$

and $g$ is the log-likelihood ratio

$$g(X_{i1}, X_{i2}, \ldots, X_{iN}) = \sum_{n=1}^{N} \left[ \frac{\mu_n - 1}{\mu_n} X_{in} - \log \mu_n \right]. \tag{18}$$

This nonlinearity $g$ can not be implemented in this situation because the parameters $\{\mu_n\}$ are not known. Many nonlinearities can be considered instead [see Appendix for example], but we will investigate the following linear "nonlinearity"

$$g(X_{i1}, X_{i2}, \ldots, X_{iN}) = \sum_{n=1}^{N} [X_{in} - 1 - \delta] \tag{19}$$

where $\delta$ is a parameter which is strictly positive. The linearity $x - 1$ is the locally optimal nonlinearity for testing between the hypotheses "$\mu = 1$" and "$\mu > 1$" for univariate exponential distributions as shown below:

$$\begin{aligned} g_{lo}(x) &= \lim_{\mu_1 \downarrow 1} \frac{d}{d\mu} \log \frac{f_\mu(x|\mu = \mu_1)}{f_\mu(x|\mu = 1)} \\ &= \lim_{\mu_1 \downarrow 1} \frac{d}{d\mu} \log \frac{\frac{1}{\mu_1} \exp\{-x/\mu_1\}}{\exp\{-x\}} \\ &= \lim_{\mu_1 \downarrow 1} \frac{d}{d\mu} \{-\log \mu_1 - x/\mu_1 + x\} \\ &= \lim_{\mu_1 \downarrow 1} x/\mu_1^2 - 1/\mu_1 \\ &= x - 1. \end{aligned}$$

Moreover, this nonlinearity is uniformly most powerful for all $\mu > 1$ in the same univariate hypothesis testing problem. In the multiparameter/multivariate case using the ESD, the

linearity given will not be the locally optimal detector because that nonlinearity requires taking the gradient in the signal direction and will result in uneven weighting of the frequency bins. This can not be done here because the $\{\mu_{in}\}$ are not known. Also, in the multiparameter case in the present context, there is no uniformly most powerful test to decide between $\mu_{in} = 1,\ n = 1, 2, \ldots, N$ and $\mu_{in} > 1,\ n = 1, 2, \ldots, N$. Despite these drawbacks, the linear function in (19) which implements an energy detection scheme should be a good test for our situation as it preserves the general structure of the optimal nonlinearity. If $\mu_{in} = \mu$ for $t_0 \leq i \leq t_1$ and $1 \leq n \leq N$, then we could make the identification from (18) and (19)

$$1 + \delta = \frac{\mu}{\mu - 1} \log \mu. \tag{20}$$

In addition to the fact that the right hand side of (20) is strictly greater than one for $\mu \neq 1$, the parameter $\delta$ is important because, as shown in Chapter 2, Page's test performs much better when the mean of the nonlinearity before the disorder takes place is negative as opposed to zero.

The false alarm rate of Page's test implemented with the nonlinearity in (19) is determined by the parameters $h$ and $\delta$. Because we have two degrees of freedom, there are many $(h, \delta)$ pairs that will yield the same false alarm rate and so one should also consider how a particular pair will affect the detectability of transients. The determination of the false alarm rate for a given pair $(h, \delta)$ does not depend on the transient but only on the noise distribution. In order to obtain the Lorden bound (4) and the Wald approximation (6) to the mean time between false alarms or equivalently the false alarm rate, it is necessary to find the root $\omega$ of the moment generating function identity (5)

$$\begin{aligned} 1 &= E\left[\exp\left\{\omega g(X_{i1}, X_{i2}, \ldots, X_{iN})\right\} \middle| \mu_{in} = 1,\ n = 1, 2, \ldots, N\right] \\ &= E\left[\exp\left\{\omega \sum_{n=1}^{N} [X_{in} - 1 - \delta]\right\} \middle| \mu_{in} = 1,\ n = 1, 2, \ldots, N\right] \end{aligned}$$

$$= \prod_{n=1}^{N} E\left[\exp\left\{\omega\left(X_{in} - 1 - \delta\right)\right\} | \mu_{in} = 1\right]$$
$$= \left(E\left[\exp\left\{\omega\left(X_{in} - 1 - \delta\right)\right\} | \mu_{in} = 1\right]\right)^{N}$$
$$= \left(\frac{1}{1-\omega}\exp\{-\omega(1+\delta)\}\right)^{N} \qquad (21).$$

Note that the root $\omega$ does not depend on $N$. (This fact may be a little surprising at first glance. One would expect that the cumulative sum would become more and more Gaussian as $N$ increased due to the Central Limit Theorem and that the root $\omega(N)$ would tend to the root for the Gaussian distribution, which is $\omega_{\text{Gauss}} = 2\delta$, as $N \rightarrow \infty$. The fallacy in this argument arises because the sum of the $N$ random variables is not normalized by $\sqrt{N}$ which would permit using the arguments that lead to the Central Limit Theorem.) After solving (21) for $\omega$ (take $N = 1$), equations (4)–(7) can be used to bound and approximate the false alarm rate.

Alternatively, one may be given the false alarm rate $\alpha$ and a minimum value $\mu_1$ of $\mu_{in}$ for which one may wish to test. The nonlinearity (19) implements the log–likelihood ratio when the identification (20) is made with $\mu = \mu_1$ and when the threshold $h'$ used with the log–likelihood ratio is related to the threshold $h$ used with (19) through

$$h = \frac{\mu_1}{\mu_1 - 1} h'. \qquad (22)$$

Then, since the root of the moment generating function identity for the log–likelihood ratio is identically equal to one, it follows that $\alpha \leq -\log h'$. Now, (22) can be used to find $h$ and (20) can be used to find $\delta$.

It is intractable to analytically evaluate the probability of detecting a transient when the transient energy can be arbitrarily distributed in the time/frequency bins. Thus, we evaluate the performance of Page's test for the following specific transient structure. For each of the blocks during the duration $\tau$ of the transient from $t_0$ through $t_1$, it is assumed

that exactly $M$ out of the $N$ bins will have transient energy, and each of the $M$ bins will have $\mu_{in} = \mu^*$ for $t_0 \leq i \leq t_1$.

The signal-to-noise ratio $S = \mu^* - 1$ required per bin to achieve a given performance measure $\tilde{\eta}$ is found from (10) and (19) to be

$$S = \frac{1}{M}\left[\frac{\tilde{\eta}}{\omega(\delta)} + N\delta\right]. \tag{23}$$

It is clear from (23) that the value of $\delta$ which minimizes the SNR does not depend on $M$.

Using the lower bound to the probability of detection in (16) we have that

$$\begin{aligned}\beta &\geq \mathcal{P}\{S_\tau > h\} \\ &= \mathcal{P}\left\{\sum_{i=1}^{\tau}\sum_{n=1}^{N}[X_{in} - 1 - \delta] > h\right\}\end{aligned} \tag{24}$$

Since typical values of $N$ are 64, 256, and 1024, the product $\tau N$ will be large enough to justify using the Central Limit Theorem to compute the probability in (24). The mean and variance of $S_\tau$ are calculated from the mean and variance of the exponential distribution yielding

$$E[S_\tau] = \tau M(\mu^* - 1) - \tau N\delta$$

$$\mathrm{Var}[S_\tau] = \tau M\mu^{*2} + \tau(N - M).$$

Thus, the bound becomes

$$\beta \geq 1 - \Phi\left(\frac{h - \tau M(\mu^* - 1) + \tau N\delta}{\sqrt{\tau\left(M\mu^{*2} + N - M\right)}}\right).$$

The signal-to-noise ratio $S = \mu^* - 1$ required for at least 50% detectability is easily found to be

$$S \leq \frac{h + \tau N\delta}{\tau M}. \tag{25}$$

## 3.5 Alternatives to Page's Test

The performance of Page's test with the linear "nonlinearity" will be compared to several other procedures. The three other tests that will be considered are the "$K$ block sliding window test," Wolcin's test, and Lorden's Test.

The "$K$ block sample sliding window test" has been called the "two models" approach in connection with the disorder problem [4] (the second "model" which is not explicitly implemented for this detector is used to estimate the background spectrum which is assumed known here.) For the present application, a fixed sample size test statistic $T_i^K$ based on the most recent $K$ blocks of $N$-point DFT's, blocks $i$ through $i+K-1$, is computed according to

$$T_i^K = \sum_{k=i}^{i+K-1} \sum_{n=1}^{N} (X_{in} - 1) \tag{26}$$

and is compared to a threshold $h_0$. The test is implemented sequentially for each new block computed by incrementing $i$ in (26) which amounts to using the new DFT and throwing away the least recent DFT. The parameter $K$ is the window length. While the probability of false alarm is easy to calculate for the fixed sample size test, it is not clear how to compute the false alarm rate for the sequential test. An approach to compute the false alarm rate might be to find the expected hitting time of a threshold by a moving average process. However, results of this type have not been found in the literature. For transients with a duration $\tau$ with $\tau \leq K$, the Central Limit Theorem can be used to lower bound the probability of detection,

$$\beta \geq 1 - \Phi\left(\frac{h_0 - \tau M(\mu^* - 1)}{\sqrt{\tau M(\mu^{*2} - 1) + KN}}\right)$$

and the SNR $S_0$ required for 50% detectability is

$$S_0 = \mu^* - 1 \leq \frac{h_0}{\tau M}. \tag{27}$$

Upon comparing this with the equivalent expression for Page's test (25), we note there is a penalty associated with Page's test due to the additional term involving $\delta$. However, the threshold for Page's test may be significantly lower than the threshold for the sliding window test for a given false alarm rate.

Wolcin's test is described in detail in the Appendix, but the relevant points are included here. For the energy detection scheme, Wolcin's test computes the following $K$ test statistics $W_k^i$ for $k = 1, 2, \ldots K$ from the most recent $K$ DFT's, numbered $i - K + 1$ through $i$,

$$W_k^i = \frac{1}{\sqrt{k}} \sum_{j=i-k+1}^{i} \sum_{n=1}^{N} (X_{jn} - 1). \tag{28}$$

For each $i$, the $K$ statistics in (28) are compared to a single threshold $h_1$ and if any exceeds the threshold, a detection is declared. Essentially, the statistics in (28) are used to detect a transient of any possible length from one to $K$. In some sense, the sliding window test is a poor-man's implementation of Wolcin's test although this is not meant to imply that Wolcin's test is necessarily better than the sliding window test. For $i$ fixed in (28), it is not clear how to compute the probability of false alarm with this multiple-hypothesis/multiple-statistic formulation. No analytical expression could be found for the false alarm rate for the sequential implementation of the test either. Using the Central Limit Theorem, the probability of detection of a transient of length $\tau$ with $\tau \leq K$ can be lower bounded by

$$\beta \geq 1 - \Phi\left(\frac{h_1 - \sqrt{\tau}M(\mu^* - 1)}{\sqrt{M\mu^{*2} + (N - M)}}\right).$$

The SNR $S_1$ required for 50% detectability is given by

$$S_1 = \mu^* - 1 \leq \frac{h_1}{\sqrt{\tau}M}. \tag{29}$$

A procedure due to Lorden [13] which extends Page's test so as to test a simple hypothesis against a composite alternative warrants investigation. Specifically, in the disorder

problem, let $F_0$ be the distribution of the i.i.d. observations before the disorder. Let $\theta$ parameterize an exponential (Koopman–Darmois) family distribution $F_\theta$ which includes $F_0$. Once the disorder takes place, the i.i.d. observations are drawn from the distribution

$$dF_\theta(y) = \exp(\theta y - b(\theta))\, dF_0(y)$$

for some $\theta \geq \theta_1$ and where $b(0) = 0$. Let $N$ be the following stopping variable:

$$N = \inf\{n : \sup_{\theta \geq \theta_1} (\theta S_n - nb(\theta)) > h_2 > 0\}, \tag{30}$$

where $S_n$ is the cumulative sum of the observations, $S_n = Y_1 + \cdots + Y_n$. Let $N^*$ be the extended stopping variable given by

$$N^* = \min_{k \geq 1} \{N_k + k - 1\} \tag{31}$$

where $N_k$ is $N$ applied to $Y_k, Y_{k+1}, \ldots$. The false alarm rate of the extended stopping rule is easily shown to be upper bounded by

$$\alpha \leq \exp\{-h\}. \tag{32}$$

Also, this procedure is optimal in the quickest detection sense asymptotically as $h \to \infty$ *for all* $\theta \geq \theta_1$. This is analogous to a uniformly most powerful test in a binary hypothesis testing problem.

The following comments are in order before using this test to detect transients from the energy spectral density. First, the optimality of the test is only in the asymptotic case whereas the optimality of Page's test in the disorder problem has been established in the non-asymptotic case as well [16]. Second, the result applies only to a single parameter exponential family distribution — this is analogous to the fact that uniformly most powerful tests exist for the single parameter exponential family distribution but not for the multiparameter distribution. Clearly, when a transient is present, the ESD has a multiparameter

exponential family distribution (17). Finally, the implementation of the test (30) requires considerably more computation and memory than any of the other tests (see [13] for details on the implementation).

Since this procedure tests for a composite alternative, one would hope that it would be robust for transients with unknown amplitude. Also, as for Page's test, the false alarm rate depends only on the null distribution of the test statistic. For the ESD, the null distribution of the test statistic has a single parameter exponential family distribution since the observations $Y_i$, $i = 1,2,\ldots$, correspond to $\sum_{n=1}^{N} X_{in}$, $i = 1,2,\ldots$ which have the Erlang–$N$ distribution,

$$p_{Y_i}(y) = \frac{y^{N-1}\exp\{-y\}}{(N-1)!}$$

for $y \geq 0$. Thus the bound (32) on the false alarm rate holds. Unfortunately, no simple bounds on the probability of detection have been obtained.

Finally, procedures which implement more than one test in parallel, such as several Page's tests in parallel or a parallel combination of Page's test and Wolcin's test, could be studied, but they will not be considered here.

## 3.6 Results

In Figure 1, the root of the moment generating function identity is plotted as a function of $\delta$ as given by (21). As is evident from (21), the root is bounded by $\omega < 1$. Since $\alpha < \exp\{-\omega h\}$, larger values of $\omega$ are desired but this goal is in opposition to having lower values of $\delta$ to have lower required SNR through (25).

Figure 2 shows the SNR required per bin for Page's test to achieve performance measures of $\tilde{\eta} = 0.1, 1$, and 10 as a function of the number of bins $M$ with transient signal energy present out of $N = 64$ and $N = 1024$ bins as given by (23). Recall, $\tilde{\eta}$ is a lower bound for the ratio $(\log T)/D$ asymptotically as the mean time between false alarms $T$ tends to infinity. Consequently, a larger $\tilde{\eta}$ corresponds to a smaller delay $D$ in detecting the disorder or transient. The values of $\delta$ used were chosen to minimize the required SNR per bin. These graphs are easily interpreted for the disorder problem as we now demonstrate. Say we want to lower bound the mean time between false alarms by $T \geq 10^8$ blocks of 64 samples each. Then, using (10) we see that for $\tilde{\eta} = 0.1$, the expected worst case delay is given by $D \approx 184.2$ blocks; for $\tilde{\eta} = 1$, $D \approx 18.4$ blocks; for $\tilde{\eta} = 10$, $D \approx 1.8$ blocks. The relations (12)-(14) can be used to yield bounds on the probability of detection of transients. For example, for large $D$, (14) can be used to show that transients of length $D$ will be detected with approximately 50% probability assuming the required SNR per bin as indicated in the figure is available.

The false alarm rates of the various tests are plotted versus the threshold in Figure 3 for $N = 64$ and $N = 1024$. Monte Carlo simulations were used to compute the false alarm rate for Page's test with $\delta = 0.1$ for $N = 64$ and with $\delta = 0.05$ for $N = 1024$. The corresponding Lorden upper bounds (4) and Wald approximations (6) are also shown. The values of $\delta$ were not chosen in any particular optimal fashion but the general guiding principle was to

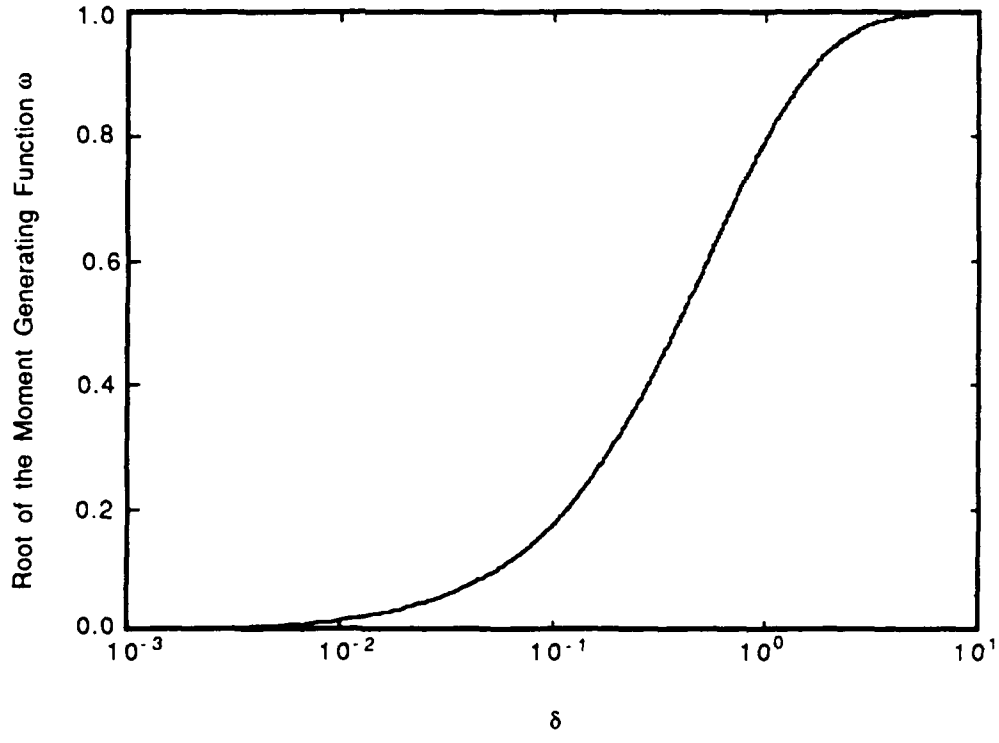
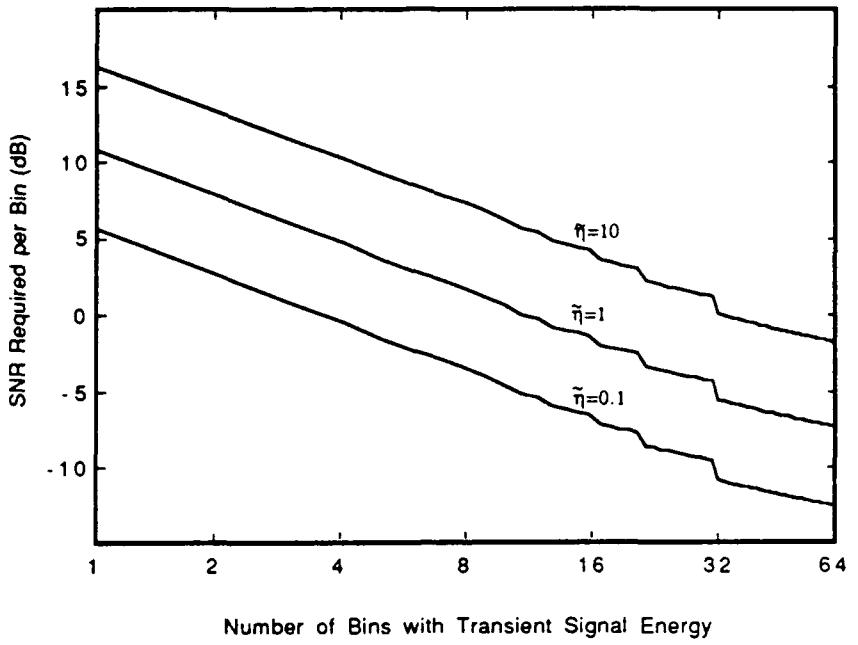
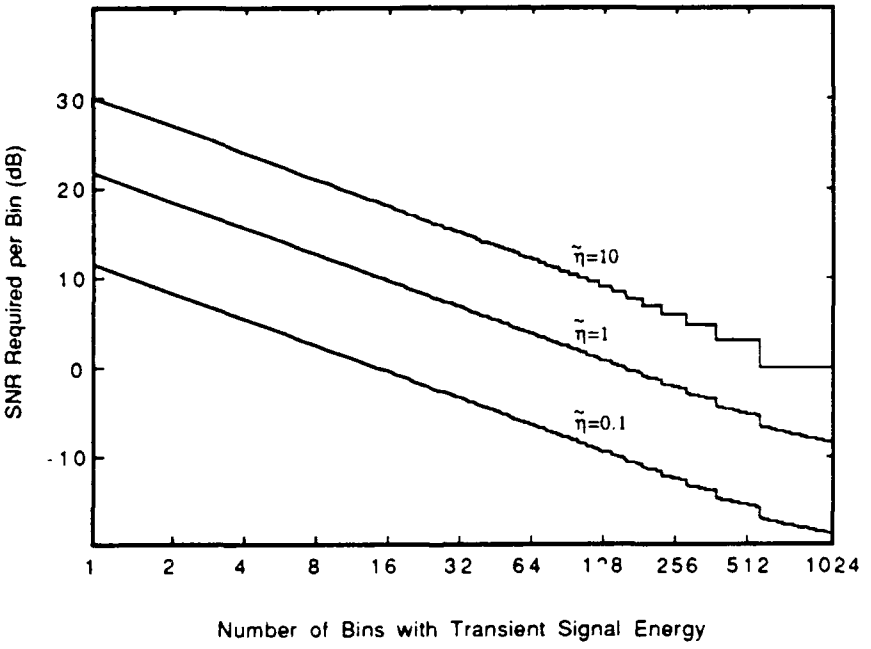



**Figure 1.** The root of the moment generating function identity given by equation (21) as a function of $\delta$.

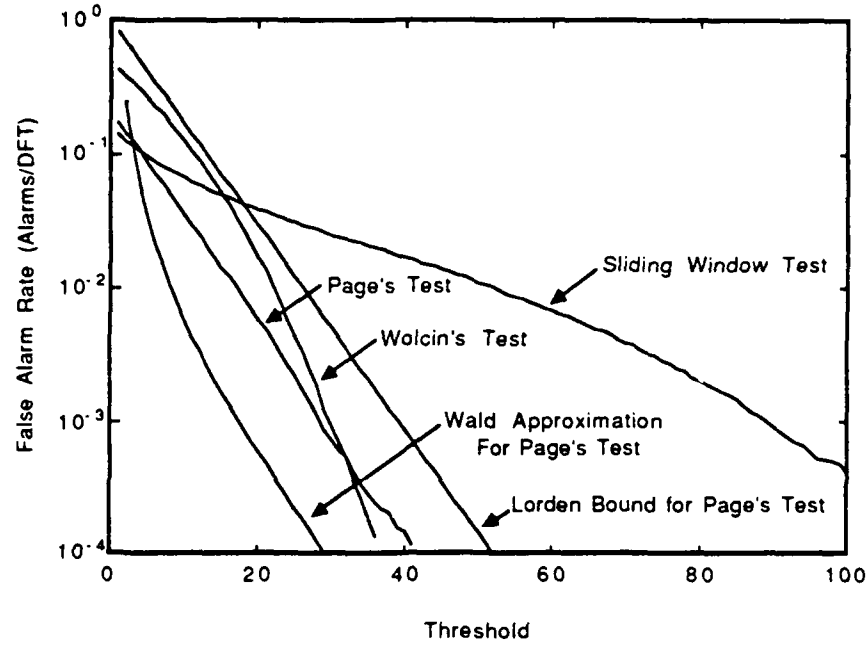



(a)

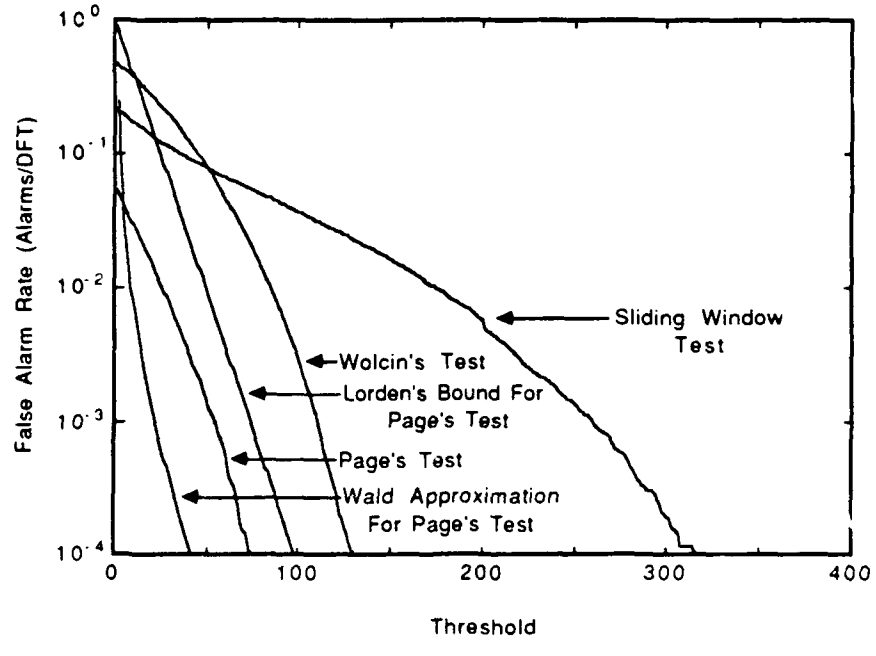



(b)

**Figure 2.** The signal-to-noise ratio required per bin for Page's test to achieve the indicated preformance $\tilde{\eta}$ as a function of the number of bins $M$ with transient signal energy present out of (a) $N = 64$ and (b) $N = 1024$ bins.




(a)

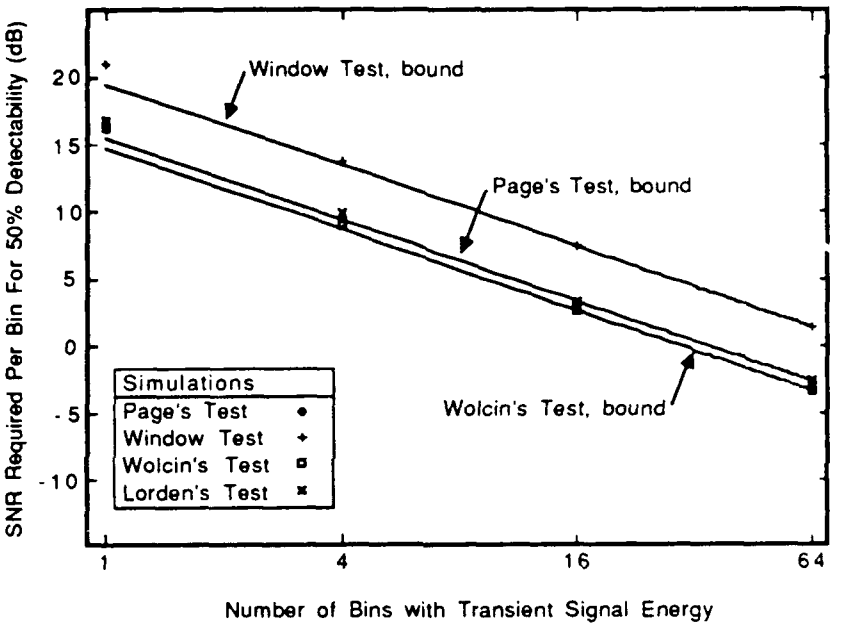



(b)

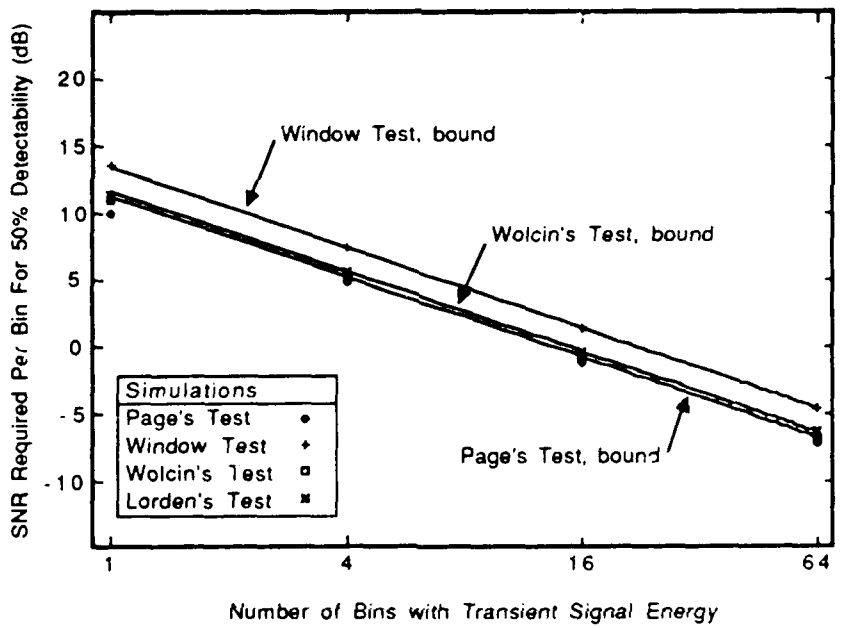
**Figure 3.** The false alarm rates of the tests considered as a function of the threshold for a DFT of length (a) $N = 64$ and (b) $N = 1024$.

lower the detection threshold by choosing a larger $\delta$ while attempting to minimize the effect that a larger $\delta$ has on the required SNR. Simulations were also run to determine the false alarm rate for the $K$-block sliding window test and Wolcin's test. The window size for both of these tests was chosen to be $K = 16$ for $N = 64$ and $K = 8$ for $N = 1024$. The upper bound for the false alarm rate for Lorden's test given by $\alpha \leq \exp(-h)$ is not shown and no simulations were performed.

The SNR required per bin for 50% detectability as a function of the number of bins with transient signal energy present, $M$, out of $N = 64$ and $N = 1024$ bins is shown in Figures 4 and 5, respectively. The false alarm rate was set at $\alpha = 0.001$ false alarms per DFT, and, with the exception of Lorden's test, the thresholds were determined from the Monte Carlo simulations depicted in Figure 3. For Lorden's test, the threshold was set using the upper bound (32), and the value of $\theta_1$ in (30) was chosen to correspond to a signal–to–noise ratio of $-20$ dB. The results are shown for transients of length $1 \leq \tau \leq 32$ when $N = 64$ and for $1 \leq \tau \leq 16$ when $N = 1024$. The upper bounds for Page's test (25), the sliding window test (27), and Wolcin's test (29) are shown together with the results of Monte Carlo simulations performed for all four tests. The simulations show that Page's test, Wolcin's test, and Lorden's test can be significantly better than the sliding window test, by approximately 5 dB, for detecting short transients and about as good to within 1 dB for detecting longer transients. Moreover, these results do not depend on the number of bins with transient signal energy, $M$. Page's test is the easiest of any of the tests to implement. It also requires the least memory, needing to store only a single scalar value. To within a half dB, it performs as well as the other tests in nearly all cases considered.

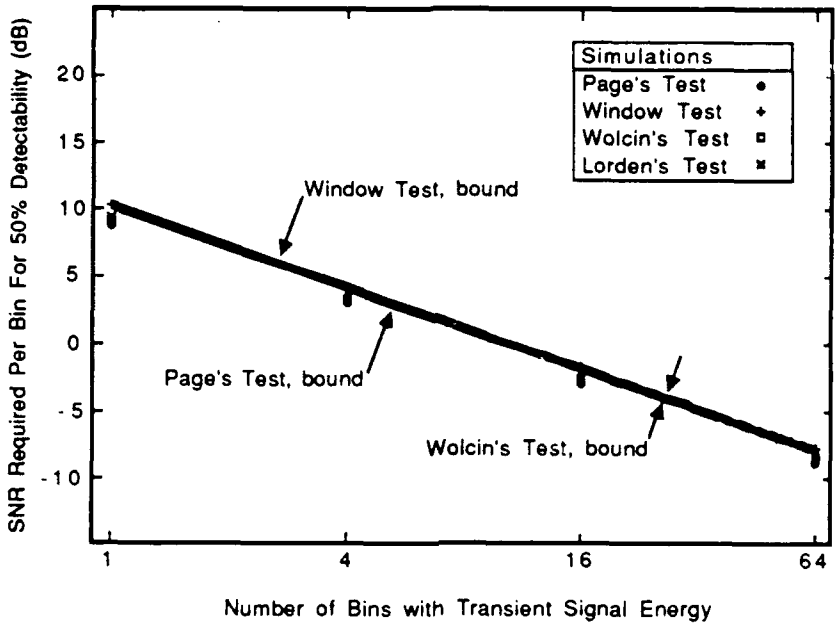



(a)

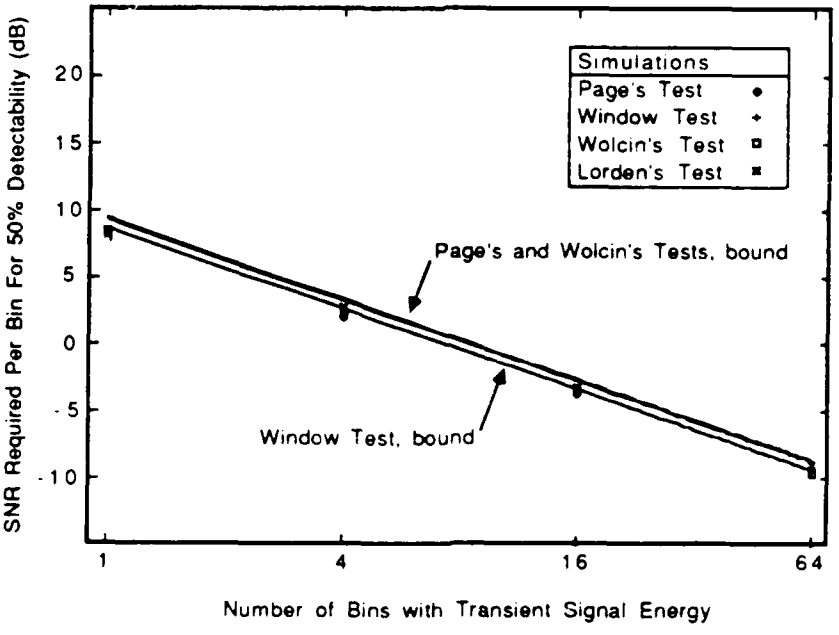



(b)

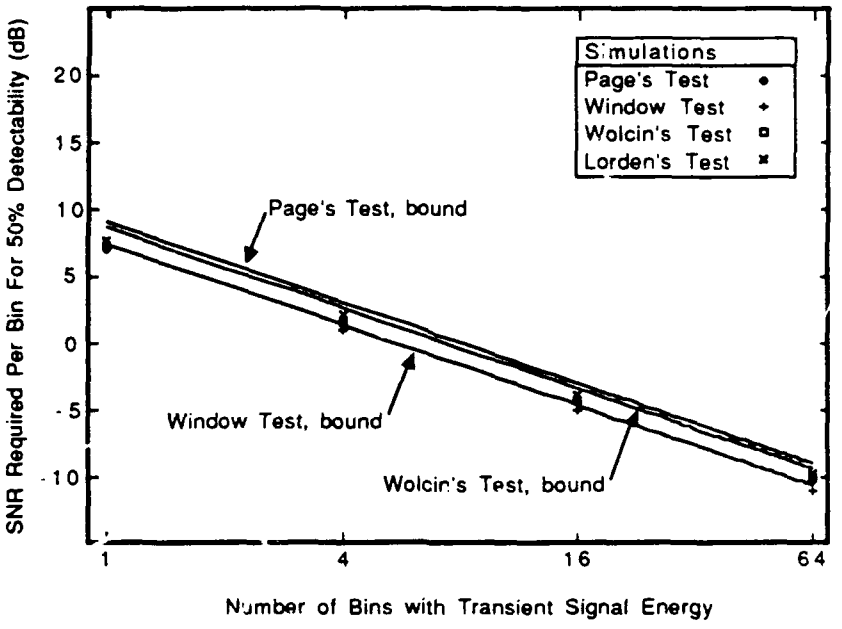
**Figure 4.** The signal-to-noise ratio required per bin for 50% detectability with false alarm rate $\alpha = 0.001$ false alarms per DFT as a function of the number of bins $M$ with signal energy for a DFT of length $N = 64$ with a transient of length of (a) $\tau = 1$, (b) $\tau = 4$, (c) $\tau = 8$, (d) $\tau = 12$, (e) $\tau = 16$, and (f) $\tau = 32$ blocks. For Page's test, $\delta = 0.1$ was used. The window size for the sliding window test and Wolcin's test is $K = 16$. Lorden's test was designed for a minimum SNR of $-20$ dB.


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
Window Test, bound
Page's Test, bound
Wolcin's Test, bound
SNR Required Per Bin For 50% Detectability (dB)
20
15
10
5
0
-5
-10
1
4
16
64
Number of Bins with Transient Signal Energy

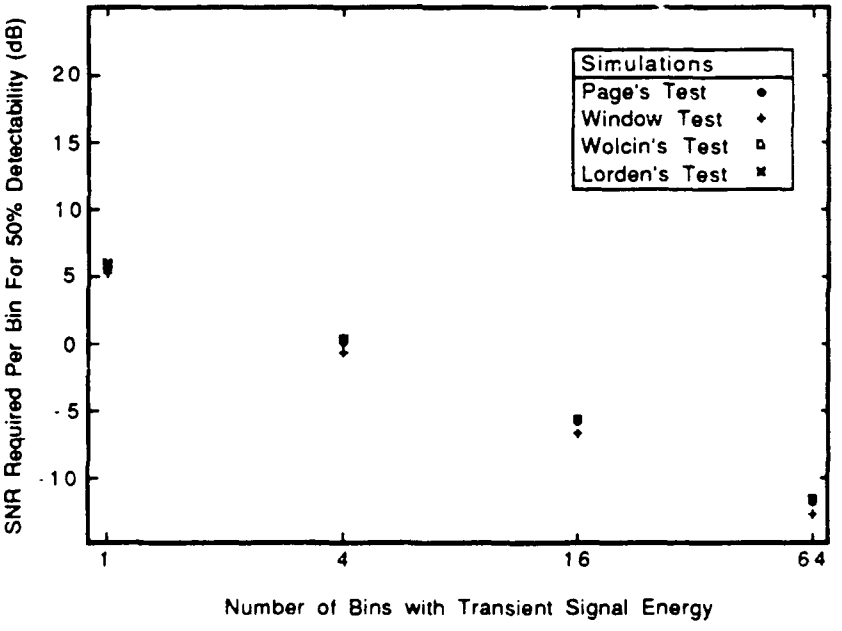
(c)


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
Page's and Wolcin's Tests, bound
Window Test, bound
SNR Required Per Bin For 50% Detectability (dB)
20
15
10
5
0
-5
-10
1
4
16
64
Number of Bins with Transient Signal Energy

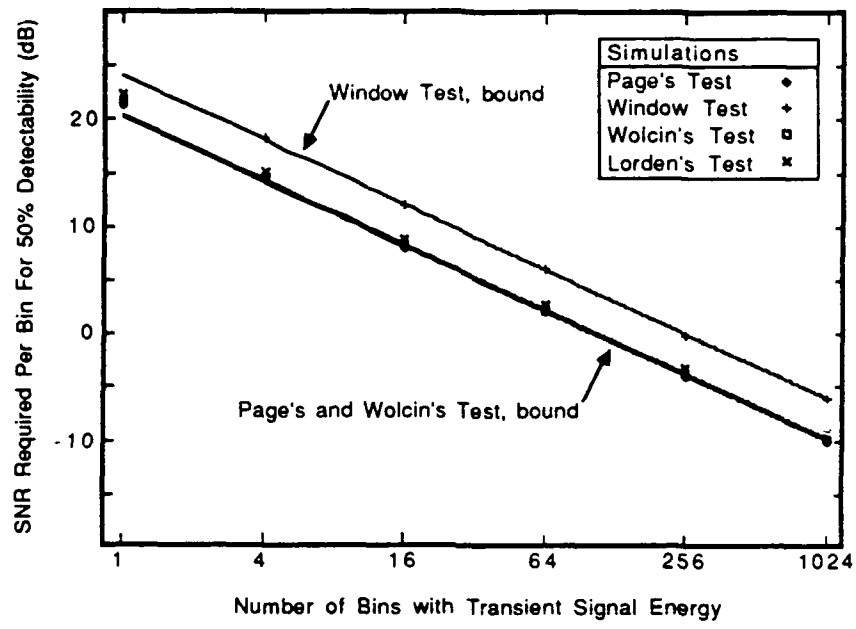
(d)


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
Page's Test, bound
Window Test, bound
Wolcin's Test, bound
SNR Required Per Bin For 50% Detectability (dB)
20
15
10
5
0
-5
-10
1
4
16
64
Number of Bins with Transient Signal Energy

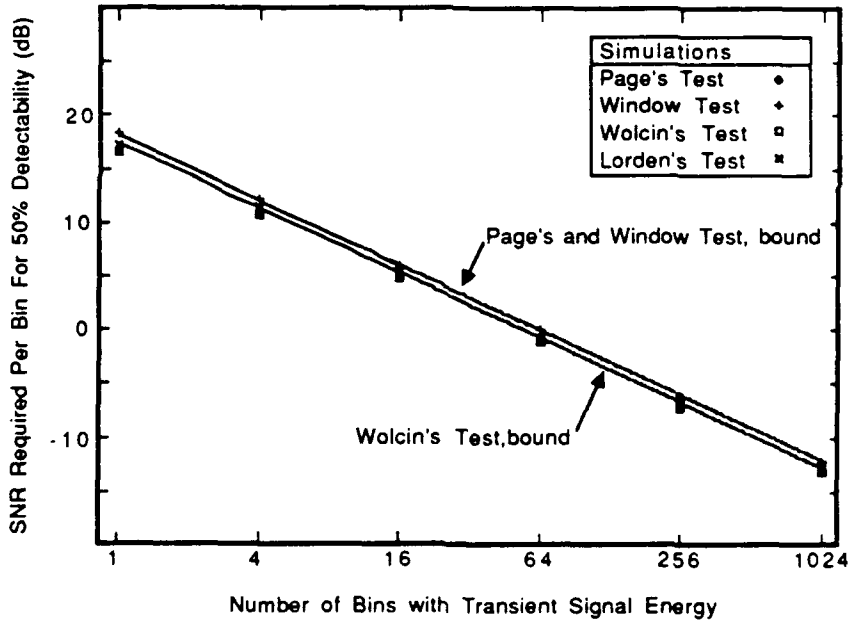
(e)


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
SNR Required Per Bin For 50% Detectability (dB)
20
15
10
5
0
-5
-10
1
4
16
64
Number of Bins with Transient Signal Energy

(f)

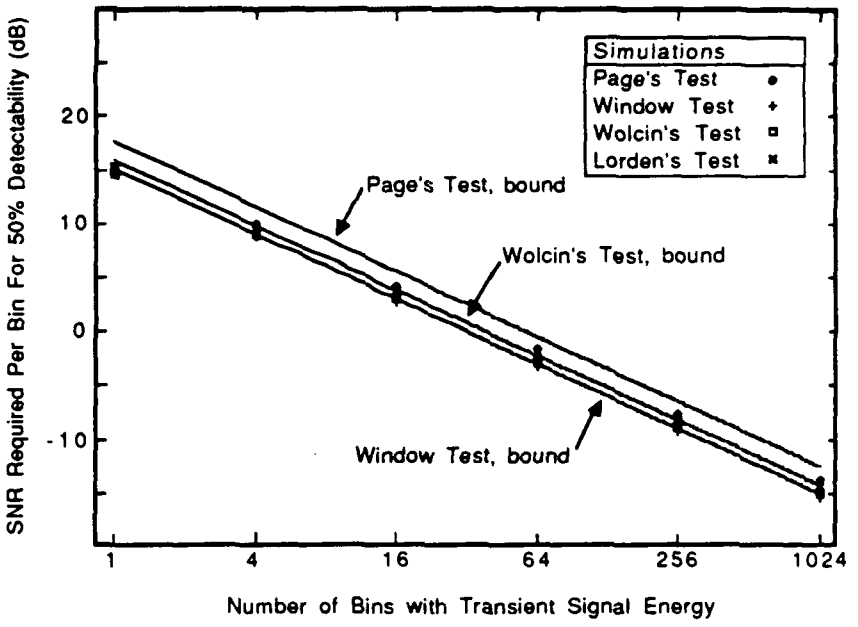



(a)

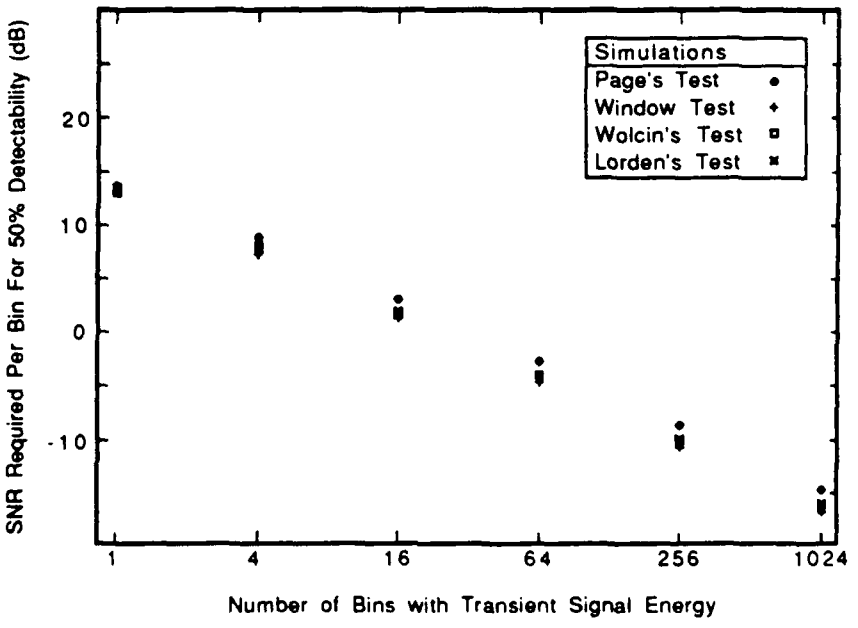



(b)

**Figure 5.** The signal-to-noise ratio required per bin for 50% detectability with false alarm rate $\alpha = 0.001$ false alarms per DFT as a function of the number of bins $M$ with signal energy for a DFT of length $N = 1024$ with a transient of length of (a) $\tau = 1$, (b) $\tau = 4$, (c) $\tau = 8$, and (d) $\tau = 16$ blocks. For Page's test, $\delta = 0.05$ was used. The window size for the sliding window test and Wolcin's test is $K = 8$ blocks. Lorden's test was designed for a minimum SNR of $-20$ dB.


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
Page's Test, bound
Wolcin's Test, bound
Window Test, bound
SNR Required Per Bin For 50% Detectability (dB)
20
10
0
-10
1
4
16
64
256
1024
Number of Bins with Transient Signal Energy

(c)


Simulations
Page's Test
Window Test
Wolcin's Test
Lorden's Test
SNR Required Per Bin For 50% Detectability (dB)
20
10
0
-10
1
4
16
64
256
1024
Number of Bins with Transient Signal Energy

(d)

## 3.7 Conclusion

The detection of transient signals is an especially difficult problem because, in general, transients occur at unknown times and have unknown durations, waveforms, and amplitudes. By using maximum likelihood considerations we have addressed the problems of unknown arrival time and unknown duration of transient signals, and the resulting detector, Page's test, performs favorably in this regard. Various detectors which have been proposed in the literature were analyzed in the context of transient detection and simulations were used to determine their performance. Page's test, which is the easiest of the tests to implement, performed very well in comparison to the other tests.

## 3.8 Appendix – Summary of the Approach of Wolcin

### 3.8.1 Introduction

In [23], Wolcin derived a test for detecting transient signals using the sample energy spectral density as the observation. Fixed sample size (FSS) tests were considered where the test operates on the samples in a $K \times N$ time/frequency grid. The design objective of the detector was that it should be robust with respect to the distribution of the transient signal energy in the two dimensional time–frequency space. Moreover, it should not assume a particular model for the transient signal, such as an autoregressive moving average model (ARMA). Wolcin derived a generalized likelihood ratio detector and showed that in certain situations this approach outperformed simple energy detection by requiring a lower SNR for 50% detectability. By combining the GLR detector with energy detection, Wolcin proposed to achieve good performance in all or most situations. Because the time when the transient occurs is not known, Wolcin uses a sequential algorithm which successively performs multiple FSS tests. We now review Wolcin's procedure in some detail.

### 3.8.2 The Robustness Objective

As stated above, a detector is desired which will perform well for a wide variety of transient signals. No structure is assumed for the process causing the transient. The transient can be narrowband in nature such as a sinusoid with finite duration or an exponentially damped sinusoid, or the transient can be broadband. Since it is assumed that the transient signal is independent of the background noise, the presence of a transient results in an increase in the total energy (power) which is to be detected by processing the observations of the sample energy spectral density. The actual increase in energy in a particular

time/frequency bin due to the presence of a transient signal is assumed unknown, further complicating the detection procedure.

### 3.8.3 Unknown Background Energy Spectral Density

In addition to not knowing the energy spectral density of the transient, the ESD of the background noise is assumed unknown, as well. The background ESD is assumed to be stationary or slowly varying with time so that the following spectral estimation can be implemented. The time samples are grouped in blocks of $N$ for processing via the Discrete Fourier Transform (DFT). Let $X_{kn}$ denote the squared magnitude of the output of the DFT for the $k$'th time block and $n$'th frequency bin. Wolcin uses the recursive estimate for the unknown ESD

$$\hat{G}_{kn} = (1 - \alpha)\hat{G}_{k-1,n} + \alpha Z_{kn}, \tag{33}$$

where $\hat{G}_{kn}$ is the estimate of ESD for the $n$'th frequency bin at time $k$ and where $\alpha$ is a parameter between zero and one. This estimate gives past observations an exponentially decreasing weight and this allows the estimate to change slowly with time. The parameter $\alpha$ is chosen to compromise the objectives of speed in adapting to changes in the background spectrum versus the steady-state variance of the estimate. After initializing the estimate and allowing it to stabilize, the spectral samples can be normalized or whitened by the spectral estimate by

$$X_{kn} = Z_{kn}/\hat{G}_{kn}.$$

If a transient signal was detected in a particular time/frequency bin, then that sample should not be used to estimate the background noise level so that (33) is modified to

$$\hat{G}_{kn} = \begin{cases} \hat{G}_{k-1,n}, & \text{if transient detected in bin } (k,n) \\ (1 - \alpha)\hat{G}_{k-1,n} + \alpha Z_{kn}, & \text{otherwise.} \end{cases}$$

### 3.8.4 The Fixed Sample Size Transient Detector

For a fixed number of blocks $K$ and frequency bins $N$, the detectors derived are fixed sample size detectors based on observations of the normalized energy samples $X_{kn}$ for $k = 1,2,\ldots,K$, $n = 1,2,\ldots,N$. In order to simplify the derivation of a detector, the following ideal assumptions are made about the distribution of the $\{X_{kn}\}$ (these are identical to those in Section 4). Under the null hypothesis $H_0$ when only the background noise is present, the $\{X_{kn}\}$ are assumed to be independent and identically distributed exponentially distributed random variables with a mean of one. (Note: the first and last coefficients actually have $\chi_1^2$ distributions but we assume these coefficients are not used for simplicity.) This would be the case if the input to the DFT was a white Gaussian process and if we ignored the effect that the normalization process has on the distribution of the energy samples. Under the alternative hypothesis $H_1$ where noise and transient signal are present, the energy samples are still assumed to be independent exponentially distributed random variables, but now sample $X_{kn}$ has mean $\mu_{kn}$ where $\mu_{kn} \geq 1$. Thus we have the binary hypothesis testing problem

$$H_0 : p(\mathbf{X}) = p_0(\mathbf{X}) = \exp\left(-\sum_{k=1}^{K}\sum_{n=1}^{N} X_{kn}\right), \qquad 0 < X_{kl} < \infty$$

$$H_1 : p(\mathbf{X}) = p_1(\mathbf{X}) = \prod_{k=1}^{K}\prod_{n=1}^{N} \frac{1}{\mu_{kn}} \exp\left(-X_{kn}/\mu_{kn}\right), \qquad 0 < X_{kn} < \infty.$$

Assuming the parameters $\{\mu_{kn}\}$ are known the optimal detector is given by the likelihood ratio test

$$\frac{p_1(\mathbf{X}|\underline{\mu})}{p_0(\mathbf{X})} \begin{matrix} > \\ < \end{matrix} h \quad \begin{matrix} \Rightarrow \\ \Rightarrow \end{matrix} \quad \begin{matrix} \text{Declare } H_1 \\ \text{Declare } H_0. \end{matrix}$$

Since the means $\{\mu_{kn}\}$ are unknown, they are replaced with their maximum likelihood estimates given by

$$\hat{\mu}_{kn} = \begin{cases} X_{kn}, & \text{if } X_{kn} \geq 1 + S_0 \\ 1, & \text{if } X_{kn} < 1 + S_0 \end{cases}$$

where $S$ is a parameter which we may choose to optimize. The resulting detector is a generalized likelihood ratio (GLR) test given by

$$\sum_{k=1}^{K}\sum_{n=1}^{N} Y_{kn} \underset{<}{\overset{>}{}} h \quad \begin{matrix} \Rightarrow \text{ Declare } H_1 \\ \Rightarrow \text{ Declare } H_0, \end{matrix}$$

where

$$Y_{kn} = \begin{cases} X_{kn} - 1 - \log X_{kn}, & \text{if } X_{kn} \geq 1 + S \\ 0, & \text{if } X_{kn} < 1 + S. \end{cases}$$

This detector can be compared with energy detection which is given by

$$\sum_{k=1}^{K}\sum_{n=1}^{N} X_{kn} - 1 \underset{<}{\overset{>}{}} h \quad \begin{matrix} \Rightarrow \text{ Declare } H_1 \\ \Rightarrow \text{ Declare } H_0. \end{matrix}$$

This detector implements the locally optimal detector for testing "$\mu_{kn} = \mu^* > 1, k = 1,2,\ldots,K,\ n = 1,2,\ldots,N$" versus "$\mu_{kn} = 1, k = 1,2,\ldots,K,\ n = 1,2,\ldots,N$" and is uniformly most powerful for all $\mu^* > 1$.

### 3.8.5 Performance

In order to determine the performance of the proposed detectors, it is convenient to relabel the $K \times N$ observations of $X_{kn}$ to be $\{X_n\}_{n=1}^{N'}$ where $N' = K \times N$, and similarly for $Y_{kn}$. Thus the GLR detector is

$$A = \sum_{n=1}^{N'} Y_n \underset{<}{\overset{>}{}} h \quad \begin{matrix} \Rightarrow \text{ Declare } H_1 \\ \Rightarrow \text{ Declare } H_0, \end{matrix}$$

and the energy detector is

$$B = \sum_{n=1}^{N'} X_n - 1 \underset{<}{\overset{>}{}} h \quad \begin{matrix} \Rightarrow \text{ Declare } H_1 \\ \Rightarrow \text{ Declare } H_0. \end{matrix}$$

Using a Central Limit Theorem argument for large $N'$, the distributions of the test statistics $A$ and $B$ under the two hypotheses are approximately Gaussian with the appropriate means

and variances. For the purpose of evaluating the performance, it is assumed that the transient has uniform strength $S$ in $M$ out of $N'$ bins. (This assumption is in contrast with the one made in Section 5. There, we assumed that the transient had uniform strength in $M$ out of $N$ frequency bins for each consecutive time block during the duration of the transient. They need not be the same frequency bins in each time block. Here, any $M$ out of the whole rectangular grid of $K \times N = N'$ bins can have transient energy. For example, There could be transient energy in the first and second DFT blocks, none in the third through sixth, and some in the seventh and eighth. According to the assumption of Section 5, this example would be considered as two seperate transients. However, it would be considered only one transient here. It should be noted that these differences are not critical. The assumptions were made only to make the performance analysis tractable.) Now, using the Gaussian assumption, the performance can be determined explicitly.

For the GLR detector, the mean and variance of the summand when no transient signal is present in bin $n$ are

$$\mu_0 = \mu(S_0, 0) = E_0[Y_n] = \int_{1+S_0}^{\infty} (x - 1 - \log x) \exp(-x) dx$$
$$\sigma_0^2 = \sigma^2(S_0, 0) = \mathrm{Var}_0[Y_n] = \int_{1+S_0}^{\infty} (x - 1 - \log x)^2 \exp(-x) dx - \mu_0^2,$$

and when the transient energy is present they are

$$\begin{aligned} \mu_1 = \mu(S_0, S) = E_1[Y_n] &= \int_{1+S_0}^{\infty} (x - 1 - \log x) \frac{1}{1+S} \exp\left(-\frac{x}{1+S}\right) dx \\ \sigma_1^2 = \sigma^2(S_0, S) = \mathrm{Var}_1[Y_n] &= \int_{1+S_0}^{\infty} (x - 1 - \log x)^2 \frac{1}{1+S} \exp\left(-\frac{x}{1+S}\right) dx - \mu_1^2. \end{aligned} \tag{34}$$

(Analytical expressions for $\mu(S_0, S)$ and $\sigma^2(S_0, S)$ are given in Section 3.8.6.) Thus, the detection problem becomes

$$H_0 : A \sim N(N'\mu_0, N'\sigma_0^2)$$
$$H_1 : A \sim N\left(M\mu_1 + (N' - M)\mu_0, M\sigma_1^2 + (N' - M)\sigma_0^2\right).$$

Let $\Phi(x)$ be the standard Gaussian cumulative distribution function. It is straightforward to show that the receiver operating characteristic (ROC) indicating the probability of detection

$\beta$ versus probability of false alarm $\alpha$ is

$$\beta = 1 - \Phi\left[\frac{\sqrt{N'}\sigma_0\Phi^{-1}(1-\alpha) + M(\mu_0 - \mu_1)}{\left[M\sigma_1^2 + (N'-M)\sigma_0^2\right]^{1/2}}\right]. \tag{35}$$

For energy detection, it easy to show that the detection problem becomes

$$H_0 : B \sim \mathrm{N}(0, N')$$

$$H_1 : B \sim \mathrm{N}(MS, N' + MS)$$

so that the ROC is given by

$$\beta = 1 - \Phi\left[\frac{\sqrt{N'}\Phi^{-1}(1-\alpha) - MS}{\sqrt{N' + MS}}\right]. \tag{36}$$

To illustrate the performance, Wolcin determined the signal-to-noise ratio (SNR) $S$ required per bin to acheive a 50% detection probability as a function of the number of bins $M$ with transient signal energy. In this case, by solving (35) and (36) with $\beta = 0.5$, we find that for the GLR detector

$$\mu_1 - \mu_0 = \frac{\sqrt{N'}\sigma_0\Phi^{-1}(1-\alpha)}{M},$$

and for the energy detector

$$S = \frac{\sqrt{N'}\Phi^{-1}(1-\alpha)}{M}.$$

Wolcin found that the GLR detector outperforms the energy detector for lower values of $M$ whereas the reverse is true for larger values of $M$.

### 3.8.6 Sequential Implementation

In the computer algorithm *MAXTRAN* [23] for transient detection, Wolcin modified the fixed sample size test described above for sequential implementation as follows. Each

time a new DFT is computed, say number $i$, the most recent $K$ DFT's numbered $i-K+1$ through $i$ are used to generate $K$ detection statistics $W_i^k$ for $k = 1, \ldots K$ according to

$$W_i^k = \frac{1}{\sqrt{k}} \sum_{j=i-k+1}^{i} \sum_{n=1}^{N} (Y_{jn} - \mu_0) \tag{37}$$

for the GLR detector and

$$W_i^k = \frac{1}{\sqrt{k}} \sum_{j=i-k+1}^{i} \sum_{n=1}^{N} (X_{jn} - 1) \tag{38}$$

for the energy detector. For each $i$, these $K$ statistics are compared to a single threshold $h$, and if any one of the statistics exceeds the threshold, a detection is declared. For each $k$, the statistic $W_i^k$ is in effect testing the hypothesis that a transient of length $k$ exists. The normalization in (37) and (38) by $\sqrt{k}$ ensures that when no signal is present the $K$ statistics all have the same distribution. For the GLR detector, the distribution is

$$W_i^k \sim N(0, N\sigma_0^2),$$

and for the energy detector, the distribution is

$$W_i^k \sim N(0, N)$$

independent of $k$. Now, the probability of false alarm of each individual test can be set at the same value by using a single threshold. Unfortunately, there does not seem to be a straightforward relationship between this probability of false alarm and the false alarm rate of the overall sequential test, and the performance of the sequential test was not evaluated in [23]. The probability of detection can be lower bounded by the probability that the statistic corresponding to the true length $\tau$ of the transient exceeds the threshold. For the GLR detector, that is

$$\beta \geq 1 - \Phi\left[\frac{h + (M/\sqrt{\tau})(\mu_0 - \mu_1)}{\left[(M/\tau)\sigma_1^2 + (N - M/\tau)\sigma_0^2\right]^{1/2}}\right],$$

and for the energy detection it is

$$\beta \geq 1 - \Phi\left[\frac{h - (M/\sqrt{\tau})S}{[N + (M/\tau)S]^{1/2}}\right].$$

The SNR required for 50% detectability is easily found tc be

$$\mu_1 - \mu_0 = \frac{\sqrt{\tau}h}{M}$$

for the GLR detector and

$$S = \frac{\sqrt{\tau}h}{M}$$

for the energy detector.

### 3.8.7 Analytical Formulae

Using the table of integrals in [21] and some tedious calculus yields the following expressions for $\mu(S_0, S)$ and $\sigma^2(S_0, S)$ defined in (33) which were not evaluated in [23]:

$$\mu(S_0, S) = \exp\left\{-\frac{1+S_0}{1+S}\right\}[1 + S_0 + S - \log(1+S_0)] + \mathrm{Ei}\left(-\frac{1+S_0}{1+S}\right)$$

$$\sigma^2(S_0, S) = \exp\left\{-\frac{1+S_0}{1+S}\right\}\Big[2S_0 + S_0^2 + 2S_0S + 4S + 2S^2$$
$$+ \log^2(1+S_0) - 2\log(1+S_0) - 2S_0\log(1+S_0) - 2S\log(1+S_0)\Big]$$
$$+ 2\mathrm{Ei}\left(-\frac{1+S_0}{1+S}\right)[S - \log(1+S_0)] + \left(C + \log\frac{1+S_0}{1+S}\right)^2 + \zeta(2)$$
$$+ 2\sum_{k=0}^{\infty}\frac{\left(-\frac{1+S_0}{1+S}\right)^{k+1}}{k!(k+1)^3} - \mu^2(S_0, S)$$

where

Ei is the exponential integral, $\mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{e^z}{z}dz$,

$C$ is Euler's constant, $C = 0.5772156649\ldots$,

$\zeta$ is the Riemann Zeta-function for which $\zeta(2) = \pi^2/6$.

## 3.9 References


[1] N. A. Abd Rabbo and A. Mattar, "A CUSUM test for detecting change in the transfer functions of open loop stochastic systems," *J. Comput. Appl. Math.*, vol. 4, no. 2, pp. 147–153, 1978.

[2] R. Andre-Obrecht, "On line segmentation of speech signals without prior recognition," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, eds., Springer-Verlag, New York, pp. 314–354, 1986.

[3] M. Basseville, "Edge detection using sequential methods for change in level – Part II: Sequential detection of change in mean," *IEEE Trans. Acous., Speech, Signal Processing*, vol. ASSP-29, no. 1, pp. 32–50, Feb. 1981.

[4] M. Basseville and A. Benveniste, "Sequential detection of abrupt changes in spectral characteristics of digital signals," *IEEE Trans. Information Theory*, vol. IT-29, no. 5, pp. 709–724, Sept. 1983.

[5] A. Benveniste, "Advanced methods of change detection: An overview," in *Detection of Abrupt Changes in Signals and Dynamical Systems*, M. Basseville and A. Benveniste, eds., Springer-Verlag, New York, pp. 77–99, 1986.

[6] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. II, John Wiley & Sons, New York, 1966.

[7] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acous., Speech, Signal Processing*, vol. 37, no. 2, pp. 169–180, Feb. 1989.

[8] S. M. Kay and L. L. Scharf, "Invariant detection of transient ARMA signals with unknown initial conditions," *Proc. IEEE Int. Conf. Acous., Speech, Signal Processing*, San Diego, CA, Mar. 19–21, 1984, paper 38.4.

[9] R. A. Khan, "Some first passage problems related to cusum procedures," *Stoch. Proc. Appl.*, vol. 9, pp. 207–215, 1979.

[10] ——, "A note on Page's two-sided cumulative sum procedure," *Biometrika*, vol. 68, no. 3, pp. 717–719, 1981.

[11] H. T. Le, "A functional analytic approach to transient signal detection and estimation," *Center for Computational Statistics and Probability*, George Mason University, Fairfax VA, Technical Report No. 45, May 1989.

[12] H. T. Le and E. J. Wegman, "Transient signal detection and estimation: a survey," *Center for Computational Statistics and Probability*, George Mason University, Fairfax VA, Technical Report No. 31, Aug. 1988.

[13] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 1897–1908, 1971.

[14] ——, "Open-ended tests for Koopman–Darmois Families," *Ann. Statist.*, vol. 1, no. 4, pp. 633–643, 1973.

[15] S. L. Marple Jr., "Transient signal enhancement by linear predictive eigenanalysis," *Proc. IEEE Int. Conf. Acous., Speech, Signal Processing*, New York, N.Y., Apr. 11–14, 1988, paper U2.8, pp. 2925–2928.

[16] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, no. 4, pp. 1379–1387, 1986.

[17] I. V. Nikiforov, "Modification and analysis of the cumulative sum procedure," *Automation and Remote Control*, vol. 41, no. 9, part I, pp. 1247–1252, 1980.

[18] B. Porat and B. Friedlander, "Adaptive detection of transient signals," *IEEE Trans. Acous., Speech, Signal Processing*, vol. 34, no. 6, pp. 1410–1418, Dec. 1986.

[19] M. B. Priestly, "Evolutionary spectra and non-stationary processes," *J. Royal Statist. Soc. B*, vol. 27, no. 2, pp. 204–237, 1965.

[20] ——, *Spectral Analysis and Time Series*, Academic Press, New York, 1981.

[21] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series: Vol. 1: Elementary Functions*, Gordon and Breach Science Pub., New York, p. 249, 1986.

[22] T. Subba Rao, "A cumulative sum test for detecting change in time series," *Int. J. Control*, vol. 34, no. 2, pp. 285–293, 1981.

[23] J. J. Wolcin, "Maximum likelihood detection of transient signals using sequenced short-time power spectra," TM 831138, Naval Underwater Systems Center, Aug. 1983.

# — 4 —

# The Gabor Representation

## 4.1 Introduction

Classical time domain analysis and frequency domain analysis are the cornerstones of most of the signal processing work in electrical engineering. These approaches, however, can be inadequate in a variety of signal processing situations as the following example demonstrates. Consider a single musical note played on some instrument, say a flute, for a few seconds. The human ear hears only a single frequency being played and yet classical Fourier theory implies that this is not the case. It is well known that any time-limited signal has a Fourier transform or frequency spectrum with infinite support. It is apparent that the classical theory, despite its unimpeachable validity, is inadequate for describing human observation and perception. This inadequacy was addressed by developing hybrid representations for signals which combine the temporal and spectral aspects of a signal, in contrast with classical analysis which deals with only one of the aspects at a time. Approaches that will not be discussed here are the "instantaneous power spectra" [13] and the "evolutionary spectra" [17] both of which define a frequency spectrum that is localized at a particular time and changes with time. Other hybrid representations are the Wigner

distribution and the (cross) ambiguity function [12]. Another approach was explored in the seminal paper by Gabor in 1946 [5].

Gabor proposed representing a signal $s(t)$ according to the expansion

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{mn} g(t - \alpha n) \exp[j2\pi\beta m(t - \alpha n)] \qquad (1)$$

where $\{C_{mn}\}$ are complex coefficients, $g(t)$ is the window function with unit energy, and $\alpha$ and $\beta$ are real parameters. Gabor originally considered a Gaussian-shaped window function and this will be considered in more detail in Section 3. The key feature of the representation (1) is that the coefficients correspond to a lattice of points in the two-dimensional space corresponding to time and frequency. For example, by using a rectangular pulse as the window function, the musical note which was inadequately described by classical representations can be represented exactly in the Gabor representation by one nonzero coefficient for a particular arrival time and frequency, and *this representation coincides with human* perception. Because the Gabor representation results in a parsimonious representation of a continuous signal that combines temporal and spectral aspects, it has recently been the subject of further study in various applications. The representation was first applied by Gabor to the analysis of hearing [5], and has more recently been used in connection with image representation and vision analysis [15,16], and with transient signal analysis [4].

Throughout this chapter, the signal $s(t)$ is assumed to be sufficiently well behaved so as to admit a Gabor representation. The existence and uniqueness of the Gabor coefficients has been investigated by various authors and references can be found in [7]. The representation is least problematic when $\alpha\beta = 1$ and, as has been done frequently throughout the literature, this condition is adopted here, too. Since the time axis in (1) can be scaled according to $t' = t/\alpha$ and the window function can be scaled according to $g'(t) = \sqrt{1/\beta}\, g(t/\beta)$, it suffices

to consider the case $\alpha = \beta = 1$. Now, the representation is simplified to

$$s(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{mn} g(t-n) \exp(j2m\pi t). \tag{2}$$

Signals with forms similar to (1) or (2) have been considered in other contexts outside the signal representation framework [19]. The signal in (2) could model a multiuser environment with a single channel where each user is assigned a frequency index $m$ and each user transmits a bit sequence modulated by the window function. Alternatively, the window function can represent the channel spreading which causes intersymbol interference (ISI). In the multiuser context, the symbol alphabet of the coefficients $\{C_{mn}\}$ is finite as in digital communication where $C_{mn}$ can take only the values $+1$ and $-1$. However, in the context of the Gabor representation, the coefficients can be arbitrary complex numbers. At first, this might seem a complication of the multiuser situation, but actually it is a simplification. Now, any optimization done over the coefficients is an unconstrained optimization and this will result in analytically tractable solutions, in contrast with solutions which implement the Viterbi algorithm in the multiuser/ISI setting.

In Section 2, the determination of the Gabor coefficients is discussed and two methods are compared when the desired signal is corrupted by noise. In Section 3, the representation is considered when a Gaussian–shaped window function is used, which is the most commonly used window function in the literature. In Section 4, an exponential window function is considered and the representation is applied to transient signal detection for finite observation intervals. The results of this section are extended to infinite observation intervals in Section 5, resulting in a sequential detection algorithm for transient signal detection.

## 4.2 The Gabor Coefficients

### 4.2.1 Introduction

The transform pairs, the forward transform and the inverse transform, related to the various Fourier representations — the Fourier transform, the Fourier series, and the discrete Fourier transform — are used frequently in the signal processing community. The special property that makes these transform pairs tractable is the orthogonality of sinusoids. As is obvious from the Gabor representation (2), the basis functions are not orthogonal for general window functions. Consequently, the forward transform to compute the Gabor coefficients is considerably more complicated than in the more familiar transforms.

### 4.2.2 The Wigner Transform

Before considering the methods for the forward transform, it is instructive to look at the generalized Wigner transform, also called the complex spectrogram. Consider a window function $g(t)$ of unit energy and define the generalized Wigner transform $S(\tau,\omega)$ of a signal $s(t)$ according to [1,6,12]

$$S(\tau, f) = \int_{-\infty}^{\infty} s(t)g^*(t-\tau)\exp(-j2\pi ft)dt, \tag{3}$$

where $g^*(t)$ is the complex conjugate of $g(t)$. Then, the inverse transform is given by [see 12 for details]

$$s(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(\tau, f)g(t-\tau)\exp(j2\pi ft)d\tau df. \tag{4}$$

Discretizing the double integral in (4) by using a rectangular approximation and sampling the integrand at $\tau = 0, \pm1, \pm2, \ldots$ and at $f = 0, \pm1, \pm2, \ldots$ results in the following approximation

$$s(t) \approx \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} S(m,n)g(t-n)\exp(j2m\pi t). \tag{5}$$

Except for the approximation, (5) is exactly the Gabor representation in (2). This suggests

similarly discretizing the integral in (3) to approximate the coefficients as follows

$$\begin{aligned} S(m,n) &\approx \sum_{k=-\infty}^{\infty} s(k)g^*(k-n)\exp(-j2\pi mk) \\ &= \sum_{k=-\infty}^{\infty} s(k)g^*(k-n). \end{aligned} \tag{6}$$

Unfortunately, the approximation is poor and is not even a function of the frequency index $m$. Nevertheless, it is apparent from (5) that the Gabor representation and the Wigner transform are related. These relationships are explored in [1,2,7].

### 4.2.3 The Gabor Coefficients via the Biorthogonal Function

There are two approaches to computing the Gabor coefficients. One has been used widely in the literature while the other has been curiously overlooked in the work on the Gabor representation, despite the fact that the tools and techniques have been available for quite some time. The popular approach is presented first. Details can be found in [1,2,3,4,7].

For a signal $x(t)$, define its Zak transform $(\mathcal{Z}x)(z,f)$ according to

$$(\mathcal{Z}x)(z,f) = \sum_{k=-\infty}^{\infty} x(z-k)\exp(-j2k\pi f).$$

Then, the Gabor coefficients for $s(t)$ can be computed from

$$C_{mn} = \int_0^1 \int_0^1 \frac{(\mathcal{Z}s)(z,f)}{(\mathcal{Z}g)(z,f)} \exp[-j2\pi(mz+nf)]dzdf.$$

A more convenient formula is obtained by considering time domain processing instead of Zak transform domain processing as follows. Define the function $\gamma(t)$ according to

$$\gamma(t) = \int_0^1 \frac{df}{(\mathcal{Z}g)^*(t,f)},$$

assuming this integral exists. The function $\gamma(t)$ is called the biorthogonal function of $g(t)$ because it exhibits the property

$$\int_{-\infty}^{\infty} g(t-n)\gamma^*(t-l)\exp[j2\pi(m-k)t]dt = \delta(n-l)\delta(m-k). \tag{7}$$

Using this property, the Gabor coefficients can easily be seen to be given by

$$C_{mn} = \int_{-\infty}^{\infty} s(t)\gamma^*(t-n)\exp(-j2\pi mt)dt. \tag{8}$$

This method of computing the coefficients will be called the *biorthogonal function method* (BFM).

### 4.2.4 The Gabor Coefficients via the Projection Theorem

The second approach to computing the Gabor coefficients can be developed in two ways, one using a Hilbert space approach and one using a random variable approach. We consider the Hilbert space development first.

It will be assumed that the functions $\{f_{mn},\ m,n \in \mathbb{Z}\}$ are linearly independent over the observation interval $I \subset \mathbb{R}$ where

$$f_{mn}(t) = g(t-n)\exp(j2m\pi t). \tag{9}$$

If the biorthogonal function exists, then the following lemma establishes the linear independence of the functions when defined over $\mathbb{R}$.

**Lemma.** *If the biorthogonal function $\gamma(t)$ corresponding to $g(t)$ exists, then the set of functions defined by* (9) *are linearly independent over* $\mathbb{R}$.

**Proof.** Assume that the functions are linearly dependent. This implies that for some $k$ and $l$ and coefficients $\{C_{mn}\}$

$$f_{kl}(t) = \sum_{m\neq k}\sum_{n\neq l} C_{mn}f_{mn}(t)$$

or equivalently

$$g(t-l)\exp(j2k\pi t) = \sum_{m\neq k}\sum_{n\neq l} C_{mn}g(t-n)\exp(j2m\pi t) \tag{10}$$

for all $t \in \mathbb{R}$. Multiplying both sides of (10) by $\gamma^*(t-l)\exp(-j2k\pi t)$ and integrating over $\mathbb{R}$ yields

$$\int_{-\infty}^{\infty} g(t-l)\gamma^*(t-l)dt = \sum_{m\neq k}\sum_{n\neq l} C_{mn} g(t-n) \int_{-\infty}^{\infty} g(t-n)\gamma^*(t-l)\exp[j2(m-k)\pi t]dt. \quad (11)$$

By the biorthogonality property (7), the left hand side of (11) is equal to 1 while the right hand side is equal to 0. Thus, the functions must be linearly independent. ■

Let $g(t)$ be a function in $L_2(I)$, the Hilbert space given by

$$L_2(I) = \{f : I \to \mathbb{C};\ \|f\|_{L_2} < \infty\}$$

where the norm $\|\cdot\|_{L_2}$ is induced by the inner product

$$\langle f, g\rangle = \int_I f(t)g^*(t)dt, \quad (12)$$

so that $\|f\|_{L_2} = \langle f, f\rangle^{1/2}$. Consider the space of functions $G(I)$ defined on $I$ given by

$$G(I) = \{f \in L_2(I) : f = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{mn} g(t-n)\exp(j2m\pi t) \text{ for some } \{C_{mn}\} \in \mathbb{C}\}.$$

It is apparent from its definition that $G(I)$ is a linear vector space. Since $G(I) \subset L_2(I)$, the inner product on $L_2(I)$ in (12) is also an inner product on $G(I)$. Thus $G(I)$ is a pre–Hilbert space. In order to show that $G(I)$ is a Hilbert space, $G(I)$ must be shown to be complete, that is, any Cauchy sequence in $G(I)$ must converge to an element in $G(I)$. This is demonstrated by the following theorem.

**Theorem.** *$G(I)$ is Hilbert space.*

**Proof.** From the discussion above it suffices to show that $G(I)$ is complete. First, the lattice of points $L$ given by

$$L = \{(m, n) : m, n \in \mathbb{Z}\}$$

is linearized and ordered as depicted in Figure 1. Now, instead of the functions $\{f_{mn};\ m,n \in \mathbb{Z}\}$ in (9), we have $\{\tilde{f}_n;\ n \in \mathbb{N}\}$.

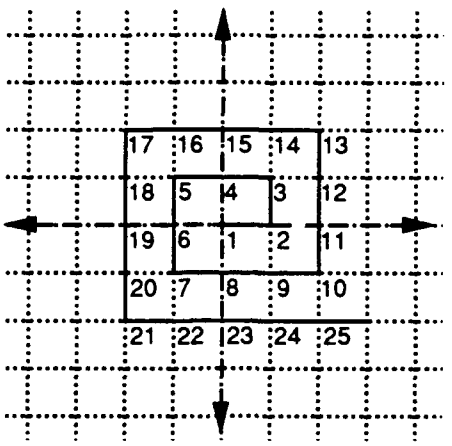



**Figure 1.** An ordering of the points in $L$.

By the Gram–Schmidt orthogonalization procedure [10], a countable set of orthonormal functions $\{e_n;\ n \in \mathbb{N}\}$ is obtained satisfying for each $n \in \mathbb{N}$

$$\text{span}\{\tilde{f}_1, \tilde{f}_2, \ldots, \tilde{f}_n\} = \text{span}\{e_1, e_2, \ldots, e_n\}.$$

Let $f$ be a function in $G(I)$. Using the new ordering of the lattice and the new orthonormal spanning vectors, $f$ can be written as

$$f(t) = \sum_{n=1}^{\infty} A_n e_n(t) \tag{13}$$

for $t \in I$. Since $\|f\|^2 < \infty$, it follows that $\sum_n |A_n|^2 < \infty$. In other words, $\{A_1, A_2, \ldots\}$ is in $l_2$, the Hilbert space of square summable sequences with the ordinary Euclidean inner product and norm denoted by $\|\cdot\|_{l_2}$. Let $\{f^k\}$ be a sequence in $G(I)$ so that in the form (13),

$$f^k(t) = \sum_{n=1}^{\infty} A_n^k e_n(t),$$

and let $x^k$ be the element in $l_2$ given by $x^k = \{A_1^k, A_2^k, \ldots\}$. Since

$$\|f^k - f^m\|_{L_2}^2 = \sum_{n=1}^{\infty} |A_n^k - A_n^m|^2 = \|x^k - x^m\|_{l_2}^2,$$

$\{f^k\}$ is a Cauchy sequence in $G(I)$ if and only if $\{x^k\}$ is a Cauchy sequence in $l_2$. Since $l_2$ is complete, $x^k$ converges to $x^\infty = \{A_0^\infty, A_1^\infty, \ldots\}$ in $l_2$. It now suffices to show that $f^k$ converges to the function $f^\infty$ given by

$$f^\infty(t) = \sum_{n=1}^{\infty} A_n^\infty e_n(t)$$

which is obviously in $G(I)$. Now,

$$\|f^k - f^\infty\|_{L_2}^2 = \sum_{n=1}^{\infty} |A_n^k - A_n^\infty|^2 = \|x^k - x^\infty\|_{l_2}^2 \longrightarrow 0$$

as $k \rightarrow 0$, which completes the proof. ■

Now that it has been established that $G(I)$ is a Hilbert space, the Projection Theorem and its corollaries can be exploited. Consider a finite collection of functions $\{f_1, f_2, \ldots, f_n\}$ in $G(I)$ where

$$f_i(t) = g(t - l_i)\exp(j2k_i\pi t)$$

for some $k_i$ and $l_i$, for $i = 1, 2, \ldots, n$. An arbitrary function $h \in G(I)$ can be projected onto the closed subspace $S$ generated by the collection of functions, and the result is the unique vector $\hat{h}$ in $S$ which is closest to $f$. The minimizing vector $\hat{h}$ can be expressed as

$$\hat{h}(t) = \sum_{i=1}^{n} \alpha_i f_i(t).$$

The $\{\alpha_i\}$, which are the Gabor coefficients for $\hat{h}$, can be computed from the *normal equations*:

$$\langle f_1, f_i\rangle\alpha_1 + \langle f_2, f_i\rangle\alpha_2 + \cdots + \langle f_n, f_i\rangle\alpha_n = \langle h, f_i\rangle \tag{14}$$

for $i = 1, 2, \ldots, n$. In some situations, such as when a finite observation interval is used or when frequency indices higher than some number are not allowed, it makes sense to consider

the projection of the function onto some subset of the Gabor basis functions. Moreover, when $h \in L_2(I)$ but $h \notin G(I)$, the Projection Theorem still holds. Thus, this method gives the optimal vector in the desired subspace in the sense of minimum distance, whereas the method using biorthogonal function will yield a suboptimal vector. Thus, the Hilbert space method would be preferred when $h \in G(I)$ is corrupted by noise. This will be discussed in further detail later. When all the Gabor coefficients of a function $h \in G(I)$ are desired, the normal equations are still valid but now an infinite system of linear equations has to be solved. The infinite system of equations can be solved iteratively by solving the finite set of equations and by increasing the subspace onto which the function is projected by one dimension on each iteration as the following lemma demonstrates.

**Lemma.** *Let $h \in G(I)$. Let $f_1, f_2, \ldots$ be an ordering of the Gabor functions. Let $\hat{h}_n$ be the projection of $h$ onto the subspace generated by $\{f_1, f_2, \ldots, f_n\}$. Then $\hat{h}_n \to h$.*

**Proof.** *It is easy to check that*

$$\|\hat{h}_n - h\|_{L_2}^2 = \sum_{i=n+1}^{\infty} |\alpha_i|^2 \tag{15}$$

where $\alpha_i$ is the coefficient of $e_i$ in the orthonormal representation of $h$. Since $\{\alpha_1, \alpha_2, \ldots\} \in l_2$, the right hand side of (15) tends to zero as $n \to 0$, which completes the proof. ■

Although this theorem guarantees that the finite dimensional solution $\hat{h}_n$ given by

$$\hat{h}_n(t) = \sum_{i=1}^{n} A_i^n f_i(t)$$

converges to $h$ in $G(I)$, it does not guarantee that the individual Gabor coefficients $A_i^n$ converge to the Gabor coefficients of $h$. Theorems and conditions for the convergence of the solutions of the finite systems of linear equations to the solution of the infinite system of linear equations can be found in [8].

### 4.2.5 The Gabor Coefficients via Maximum Likelihood Estimation

The method involving the normal equations can be arrived at from a random processes approach and this will further extend the optimal properties of this approach. Consider a signal $s(t) = s_t(\{C_{mn}\})$ with the Gabor representation (2) corrupted by additive complex white Gaussian noise $n(t)$ with zero mean and autocorrelation function $E[n(s)n^*(t)] = N_0\delta(t-s)$. Thus, the received signal $r(t)$ is $r(t) = s(t) + n(t)$. Given the Gabor coefficients, the likelihood of the received waveform is

$$P[\{r(t), t \in I\}|\{C_{mn};\ m,n \in \mathbb{Z}\}] = A\exp[\Omega(\{C_{mn}\})/2N_0] \tag{16}$$

where

$$\Omega(\{C_{mn}\}) = 2\Re\left\{\int_I s_t^*(\{C_{mn}\})dr_t\right\} - \int_I |s_t(\{C_{mn}\})|^2 dt, \tag{17}$$

and $A$ is a positive constant which does not depend on $\{C_{mn}\}$. Setting

$$Z_{mn} = \int_I g^*(t-n)\exp(-j2m\pi t)dr_t, \tag{18}$$

we have that

$$\int_I s_t^*(\{C_{mn}\})dr_t = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{mn}Z_{mn}. \tag{19}$$

In light of (16)-(19), it follows that the statistics $\{Z_{mn};\ m,n \in \mathbb{Z}\}$ are jointly sufficient statistics for $\{C_{mn};\ m,n \in \mathbb{Z}\}$. The second term on the right hand side of (17) is easily shown to be

$$\int_I |s_t(\{C_{mn}\})|^2 dt = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} C_{kl}\langle f_{kl}, f_{mn}\rangle C_{mn}^*. \tag{20}$$

Order the coefficients $C_{mn}$ in a vector $\mathbf{C}$, for example using Figure 1, and similarly order the statistics $Z_{mn}$ in a vector $\mathbf{Z}$. Corresponding to this ordering, define the *signal cross correlation matrix* $\mathbf{R}$ whose entries contain the inner products of the form $\langle f_{kl}, f_{mn}\rangle$. Now (17) can be rewritten as

$$\Omega(\mathbf{C}) = 2\Re\{\mathbf{C}^H\mathbf{Z}\} - \mathbf{C}^H\mathbf{R}\mathbf{C} \tag{21}$$

where $\mathbf{C}^H$ is the complex–conjugate transpose of $\mathbf{C}$. The vector $\hat{\mathbf{C}}$ which maximizes (21) is the maximum likelihood estimate of $\mathbf{C}$ and is given by the solution of the infinite system of linear equations

$$\mathbf{R}\hat{\mathbf{C}} = \mathbf{Z}, \tag{22}$$

which for $\mathbf{R}$ invertible, has solution $\hat{\mathbf{C}} = \mathbf{R}^{-1}\mathbf{Z}$. This is exactly the result obtained from the Hilbert space approach and (22) are the normal equations. If instead of considering all the Gabor coefficients only a finite number are permitted to be non-zero, then the above derivation can be used to find the MLE of $\mathbf{C}$ which will result in the finite system of normal equations in (14). This corresponds to the projection onto the finite dimensional subspace of Gabor functions. When $\mathbf{R}$ is invertible in the infinite system of linear equations, $\hat{\mathbf{C}}$ is an unbiased estimate of $\mathbf{C}$, as is demonstrated below. Since $E[Z_{mn}]$ is given by

$$\begin{aligned} E[Z_{mn}] &= E\left[\int_I g^*(t-n)\exp(-j2m\pi t)\left\{\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} C_{kl}g(t-l)\exp(j2k\pi t) + n(t)\right\}dt\right] \\ &= \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} C_{kl}\int_I g^*(t-n)\exp(-j2m\pi t)g(t-l)\exp(j2k\pi t)dt, \end{aligned}$$

it follows that $E[\mathbf{Z}] = \mathbf{RC}$. Since $E[\mathbf{C}] = \mathbf{R}^{-1}E[\mathbf{Z}]$, $E[\hat{\mathbf{C}}] = \mathbf{C}$. As done with the Projection Theorem, it can be shown that the finite dimensional approximations $\hat{s}^n(t)$ converge to $s(t)$ in the mean-square sense, but $\hat{\mathbf{C}}^n$ corresponding to the approximations need not converge in expectation to the true Gabor coefficients $\mathbf{C}$. However, if the signal is such that only a finite set of coefficients are nonzero, then the finite solution of the normal equations is unbiased, as well.

The distribution in (16) is of exponential family form and is of full rank because of the linear independence of the Gabor functions. Since $\hat{\mathbf{C}}$ is a function of the sufficient statistics and is unbiased, it follows from [9, Theorem 2.1.2] that each element in $\hat{\mathbf{C}}$ is a uniformly minimum variance unbiased estimate of the corresponding element in $\mathbf{C}$. Thus, in the presence of white Gaussian noise, the Gabor coefficients computed by the projection

method will have smaller variances than those computed with the biorthogonal function. Note that the procedure above is easily modified when the noise is non-white Gaussian noise, in which case the processing is similar to the one given by the Gauss-Markov Theorem [10]. The optimality properties of this method still hold.

The philosophical difference between the approach using the biorthogonal function and the approach using the normal equations can be clarified through the analogy with inter-symbol interference or a multiuser channel. If a single data symbol is desired in the ISI case or a single–user receiver in the multiuser channel, then one approach is to process the data so as to null out the interfering symbols or users. The other approach is to simultaneously estimate all the symbols or users and in essence "subtract out" the undesired ones from the original signal. When the interference has a structure that is known a priori, then this can be exploited in the simultaneous estimation and the second approach should be better. For signals with the Gabor representation, filtering with the biorthogonal function nulls out all the basis functions except the one of interest, but this may amplify the noise at the same time. On the other hand, the solution of the normal equations is an implementation of simultaneous estimation of the Gabor coefficients and deals with the interference, the desired signal, and the noise in a unified approach.

### 4.2.6 The Distribution of the Gabor Coefficients

In order to determine the performance of the two schemes in estimation and detection problems, it is necessary to determine the distribution of the coefficients when the schemes are applied to signals embedded in zero–mean white Gaussian noise. As both approaches process the data in a linear fashion, the sequence of estimated Gabor coefficients for white Gaussian noise calculated by either method is itself a Gaussian random process. Thus, it is sufficient to determine the first and second order moments of the process. Since both

schemes yield unbiased estimates of the Gabor coefficients for a signal with the Gabor representation embedded in zero mean noise, the means of the estimated coefficients are the true Gabor coefficients.

The second order moments are as follows. Let $n(t)$ represent the noise with autocorrelation function $E[n(s)n^*(t)] = N_0\delta(t-s)$. Then, the second order moments for the method using the biorthogonal function is easily seen to be

$$\begin{aligned} E[(\hat{C}_{kl} - C_{kl})(\hat{C}^*_{mn} - C^*_{mn})] &= N_0 \int_{-\infty}^{\infty} \gamma^*(t-l)\gamma(t-n)\exp[j2\pi(m-k)t]dt \\ &= N_0 \int_{-\infty}^{\infty} \gamma^*(t)\gamma(t-(n-l))\exp[j2\pi(m-k)t]dt. \end{aligned} \tag{23}$$

In particular, the variance of each coefficient is

$$\mathrm{Var}[\hat{C}_{mn}] = N_0 \int_{-\infty}^{\infty} |\gamma(t)|^2 dt. \tag{24}$$

When the maximum likelihood estimates for the Gabor coefficients are used, the estimates are given by $\hat{\mathbf{C}} = \mathbf{R}^{-1}\mathbf{Z}$ as given in the previous section. It is easy to check that under the white Gaussian noise assumption the vector of sufficient statistics $\mathbf{Z}$ has covariance matrix given by $E[\mathbf{Z}\mathbf{Z}^H] = N_0\mathbf{R}$. Thus, the covariance matrix of the coefficients is

$$E[(\hat{\mathbf{C}} - \mathbf{C})(\hat{\mathbf{C}} - \mathbf{C})^H] = \mathbf{R}^{-1}N_0\mathbf{R}\mathbf{R}^{-1} = N_0\mathbf{R}^{-1}.$$

### 4.2.7 Practical Computation of the Gabor Coefficients

In order to compute the Gabor coefficients by either method described above, the integrals in (8) or (18) are computed. In real applications, the integrals would be computed by approximating them with finite sums. If the biorthogonal function has infinite support,

it is necessary to truncate it to an interval, say $[-I, I]$. Then, the integral is approximated by

$$C_{mn} \approx \frac{1}{L} \sum_{l=0}^{L-1} \sum_{i=-I}^{I-1} \gamma\left(i + \frac{l}{L}\right) s\left(i + n + \frac{l}{L}\right) \exp\left(-j2\pi \frac{ml}{L}\right), \tag{25}$$

where $L$ samples are taken per second. The right hand side of (25) can be computed efficiently using the Fast Fourier Transform (FFT). Similarly, truncating the window function if necessary, the sufficient statistics in (18) are approximated by

$$Z_{mn} \approx \frac{1}{L} \sum_{l=0}^{L-1} \sum_{i=-I}^{I-1} g_G\left(i + \frac{l}{L}\right) r\left(i + n + \frac{l}{L}\right) \exp\left(-j2\pi \frac{ml}{L}\right), \tag{26}$$

which can also be computed with the FFT. The only difference between (25) and (26) is the use of the biorthogonal function in (25), while (26) uses the window function or matched filter.

### 4.2.8 The Extension to Real Signals

Both of the methods described are easily extended to real signals of the following form:

$$s(t) = \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} A_{mn} g(t - n) \cos(2\pi m t - \theta_{mn}). \tag{27}$$

If both the window function and the biorthogonal function are real valued, it can be checked that the Gabor coefficients $\{C_{mn}\}$ of a real signal such as the one in (27) will satisfy $C_{-m,n} = C_{mn}^*$. It is assumed this is so. By expanding the cosine term into complex exponentials, it is easily seen that

$$\begin{aligned}
A_{0n} \cos(\theta_{0n}) &= C_{0n} \\
A_{mn} \cos(\theta_{mn}) &= C_{mn} + C_{-m,n} = 2\Re\{C_{mn}\} \\
A_{mn} \sin(\theta_{mn}) &= (C_{mn} - C_{-m,n})/j = 2\Im\{C_{mn}\} \\
|A_{mn}|^2 &= 4|C_{mn}|^2 \\
\tan(\theta_{mn}) &= \frac{j(C_{mn} + C_{-m,n})}{C_{mn} - C_{-m,n}} = \frac{\Im\{C_{mn}\}}{\Re\{C_{mn}\}}.
\end{aligned}$$

These relationships hold for both methods discussed for complex signals operating on the real signal in (27) *in the absence of noise.*

For both methods, care must be taken when the signal (27) is corrupted by additive white Gaussian noise. It no longer makes sense to consider complex Gaussian noise since the imaginary part can be removed without loss simply by taking the real part of the signal. Thus, the noise is assumed to be real without loss of generality. When the biorthogonal function is used, this assumption does not affect the computation of the coefficients. It does, however, affect the distribution of the estimates of the coefficients. Now, in addition to the second order moments in (23), the following moments are required to fully specify the distribution:

$$E[(\hat{C}_{kl} - C_{kl})(\hat{C}_{mn} - C_{mn})] = N_0 \int_{-\infty}^{\infty} \gamma(t-l)\gamma(t-n)\exp[j2\pi(m+k)t]dt. \qquad (28)$$

The distribution of the real and imaginary parts of the Gabor coefficients can be determined from the moments in (23) and (28). See [3,4] for an example of this procedure.

For the maximum likelihood method, the complex sufficient statistics $\{Z_{mn}\}$ for the complex case are still jointly sufficient for the real signal in noise case. Equivalently, the statistics $\{\Re(Z_{mn}), \Im(Z_{mn})\}$ are a set of real statistics which are jointly sufficient. Now, however, the signal cross correlation matrix $\mathbf{R}$ must contain the real inner products

$$\int_I g(t-l)g(t-n)\cos(2\pi kt)\cos(2\pi mt)dt,$$
$$\int_I g(t-l)g(t-n)\cos(2\pi kt)\sin(2\pi mt)dt,$$
$$\int_I g(t-l)g(t-n)\sin(2\pi kt)\sin(2\pi mt)dt.$$

These correspond to the correlation between $\Re\{Z_{kl}\}$ and $\Re\{Z_{mn}\}$, between $\Re\{Z_{kl}\}$ and $\Im\{Z_{mn}\}$, and between $\Im\{Z_{kl}\}$ and $\Im\{Z_{mn}\}$, respectively. The procedure for computing

the maximum likelihood estimates proceeds as before but with the new matrix and sufficient statistics. As before, the distribution is completely determined by the signal cross-correlation matrix **R**. An example of this procedure is given in Section 4 for the exponential window function.

## 4.3 The Gaussian Window Function

The original window function that Gabor considered in his representation is the Gaussian window function given by

$$g_G(t) = \left(\frac{2}{\tau^2}\right)^{1/4} \exp[-\pi t^2/\tau^2]. \tag{29}$$

The parameter $\tau^2$ controls the width of the window and the other constants were chosen so that $\int |g(t)|^2 dt = 1$. The Gaussian window function has the property that the transform $S(f)$ of the translated and sinusoidally modulated window $s(t)$,

$$s(t) = \left(\frac{2}{\tau^2}\right)^{1/4} \exp[-\pi(t - t_0)^2/\tau^2] \exp[j(2\pi f_0 t + \phi)] \tag{30}$$

has the same form as $s(t)$, namely,

$$S(f) = (2\tau^2)^{1/4} \exp[-\pi\tau^2(f - f_0)^2] \exp\{-j[2\pi t_0(f - f_0) - \phi]\}.$$

Let $\Delta t$ be the root mean square duration of a signal $x(t)$ and let $\Delta f$ be the root mean square duration of the Fourier Transform of $x(t)$. The celebrated uncertainty principle requires that $\Delta t \Delta f \geq \pi$. Equality is achieved for the sinusoidally modulated Gaussian pulse in (30). This property implies that this signal has the maximum resolution in the time/frequency space. This was the motivation for the choice of the Gaussian window

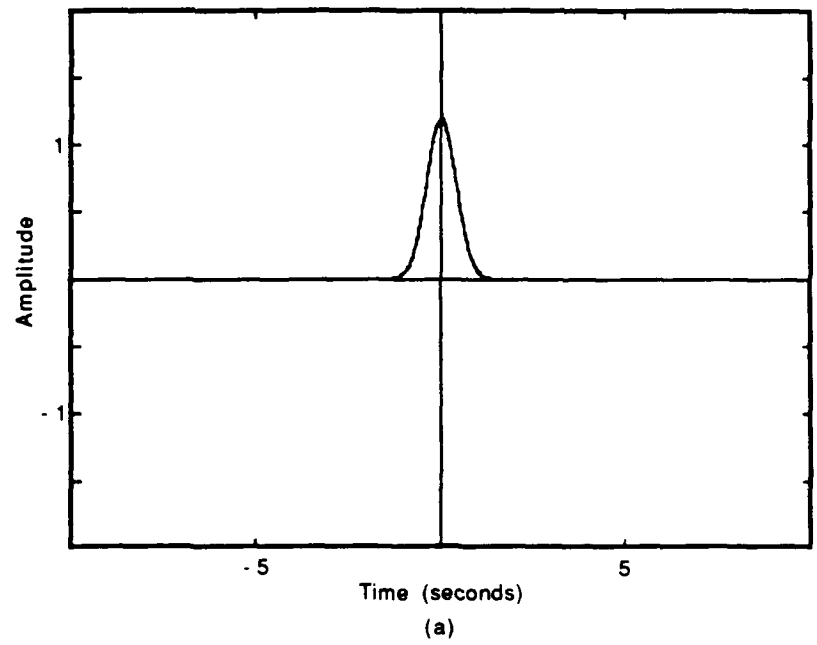
function in the Gabor representation. The Gaussian window function was later used in the two-dimensional setting of vision analysis and image representation [15,16]. Throughout the literature, the biorthogonal function of the Gaussian window in one dimension (or two for images) has been used exclusively to compute the Gabor coefficients.

The biorthogonal function for the Gaussian window was found to be [1,2]

$$\gamma_G(t) = (2\tau^2)^{-1/4} \left(\frac{K_0}{\pi}\right)^{-3/2} \exp(\pi t^2/\tau^2) \sum_{n>t/\tau-1/2} (-1)^n \exp\left[-\pi\left(n+\frac{1}{2}\right)^2\right],$$

where $K_0 = 1.85407468\ldots$ is the complete elliptic integral for the modulus $1/\sqrt{2}$. The window function and biorthogonal function are shown in Figure 2 for $\tau^2 = 1$, and in Figure 3 for $\tau^2 = 0.5$. It is clear that for smaller values of $\tau$ for which the Gaussian pulse is very narrow, the biorthogonal function concentrates its energy in areas where the Gaussian pulse is smallest in order to compute the coefficients. Therefore, in the presence of noise most of the signal energy is ignored and the noise is amplified. This is in contrast to matched filtering where the filter amplifies portions with large signal values and attenuates portions with small signal values.

As pointed out in [7], $\gamma_G(t)$ is not square integrable. It follows from (24) that if the biorthogonal function is used in the presence of white noise, the variance of the resulting Gabor coefficients would be infinite. However, in any real application the function would have to be truncated to be realizable, and since the biorthogonal function is bounded, the variance of the Gabor coefficients would be finite. Using a truncated biorthogonal function results in biased estimates of the coefficients. Thus, a balance must be struck between the estimation bias and the estimation variance when truncating the biorthogonal function.



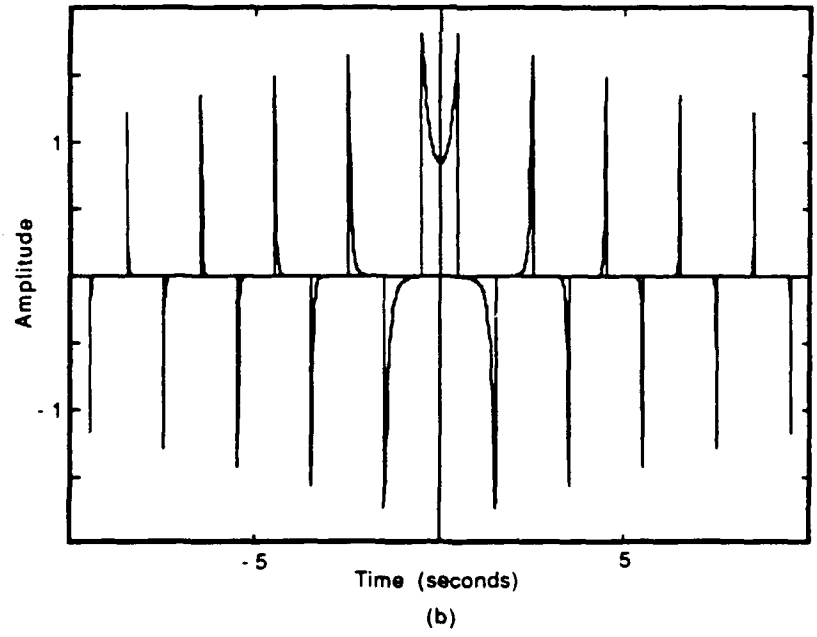




**Figure 2.** (a) The Gaussian window function $g_G(t)$ and (b) the biorthogonal window function $\gamma_G(t)$ for $\tau^2 = 1$.

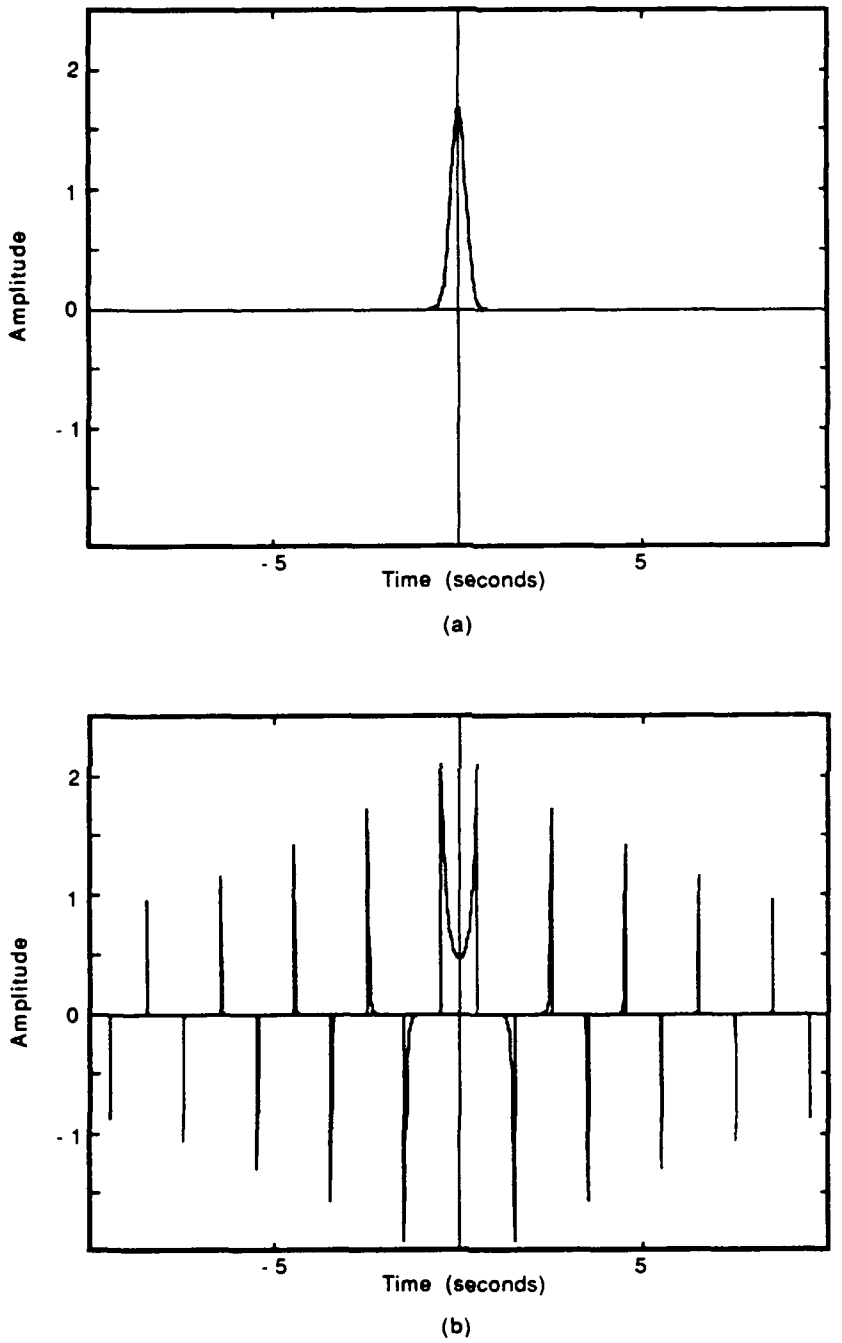



**Figure 3.** (a) The Gaussian window function $g_G(t)$ and (b) the biorthogonal window function $\gamma_G(t)$ for $\tau^2 = 0.5$.

In order to use the projection method, the signal cross correlations are required:

$$\begin{aligned}
\langle f_{kl}, f_{k+m,l+n} \rangle &= \int_{-\infty}^{\infty} g_G(t-l) g_G(t-l-n) \exp(j2\pi mt) dt \\
&= (-1)^{mn} \int_{-\infty}^{\infty} g_G(t-n/2) g_G(t+n/2) \exp(j2\pi mt) dt \\
&= (-1)^{mn} \int_{-\infty}^{\infty} \sqrt{2} \exp\left[ -\frac{\pi(t-n/2)^2}{\tau^2} - \frac{\pi(t+n/2)^2}{\tau^2} \right] \cos(2\pi mt) dt \\
&= (-1)^{mn} \exp\left( -\frac{\pi m^2 \tau^2}{2} \right) \exp\left( -\frac{\pi n^2}{2\tau^2} \right) \qquad (31) \\
&= r(m,n).
\end{aligned}$$

Note that the expression in (31) "almost" factors into a product of two terms, one which depends only on the time difference $n$ and one which depends only on the frequency difference $m$. Had the expression factored, the inversion of the signal cross correlation matrix $\mathbf{R}$ which is required for the maximum likelihood estimates would split into the product of the inversions of the two factors. If it was known, for example, that only even frequency indices $m$ can have non-zero Gabor coefficients, then (31) would factor and would result in simpler processing. This factoring would be especially useful in a sequential computation of the Gabor coefficients because of the decoupling between the delay index and the frequency index. For sequential processing, there may still be the promise of reduced computation due to the unusual form of (31), and this could be an area of study for practical implementation.

Although the linear independence of the Gabor functions insures linear independence of the columns of the signal correlation matrix, the doubly infinite matrix need not be invertible. It must be positive semidefinite though. Numerical difficulties were encountered in the inversion of finite signal cross-correlation matrices with large time and frequency dimensions, and it is hypothesized that the infinite matrix is not invertible. This does not mean that a unique solution does not exist for the normal equations. However, even

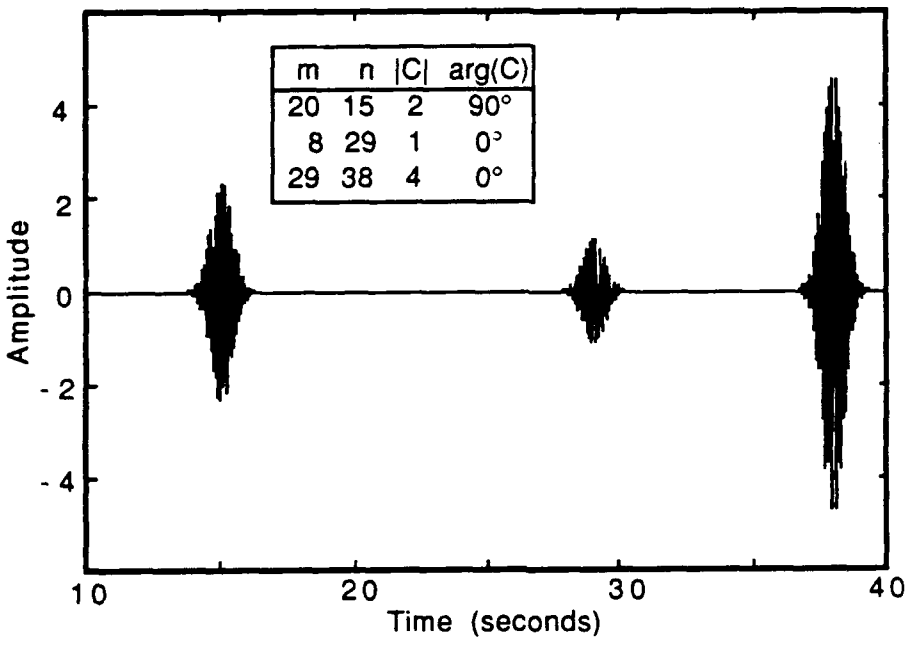
when a unique solution does exist, it would need to be computed iteratively by solving the finite system of equations for progressively higher order systems [8]. This is assumed to be impractical and so estimates can be obtained by using only the solution to a single finite system of normal equations.

Figure 4 shows the real part of a signal with the Gabor representation using the Gaussian window with $\tau^2 = 1$, and, using the same coefficients, the real part of the signal with $\tau^2 = 0.5$. The signal has three nonzero coefficients and there is no noise present. Using the biorthogonal function restricted to the interval $[-10, 10]$, the coefficients were computed by using the technique discussed in Section 4.2.7. $L = 64$ samples are taken per second. Similarly, the sufficient statistics were computed by truncating the Gaussian window to the interval $[-10 : 10]$. However, little is lost in the truncation of the Gaussian window since it decays exponentially to zero, unlike the biorthogonal function.

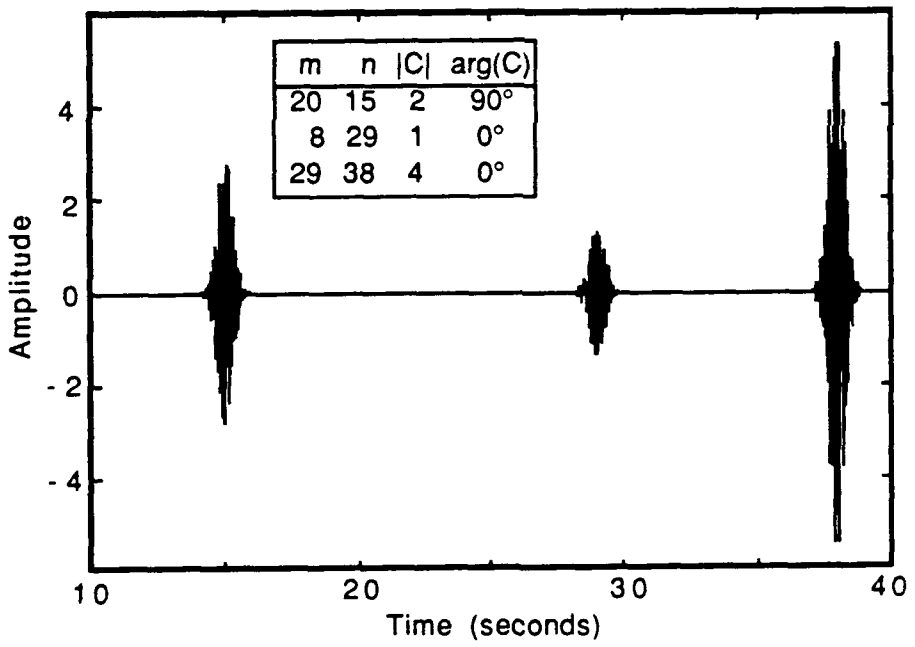
In order to compute the maximum likelihood estimates, the vector of sufficient statistics has to be multiplied by the inverse of the signal cross-correlation matrix, which is doubly infinite. Instead, a two-dimensional filter is designed to filter the sufficient statistics. This filter is obtained as follows. For a fixed maximum frequency index difference $M - 1$ and delay index difference $N - 1$, an $MN \times MN$ matrix $\hat{\mathbf{R}}$ is defined according to

$$\hat{\mathbf{R}}_{nM+m,lM+k} = r(m-k, n-l) \tag{32}$$

for $m = 0, 1, \ldots, M-1$, $n = 0, 1, \ldots, N-1$, where $r(m, n)$ is defined in (31). From the form of (32), $\hat{\mathbf{R}}$ is a Toeplitz-block, block-Toeplitz matrix. After $\hat{\mathbf{R}}^{-1}$ is computed, the middle row $\mathbf{v}^T$ of the inverse is extracted because $\mathbf{v}^T\mathbf{Z}$ is the estimate of the middle coefficient of the $MN \times MN$ grid. The elements of $\mathbf{v}$ are ordered into an $M \times N$ two-dimensional filter which is used to filter the two-dimensional array of sufficient statistics. Thus, the estimate of each Gabor coefficient is obtained from the $MN$ neighboring sufficient statistics. Since




(a)



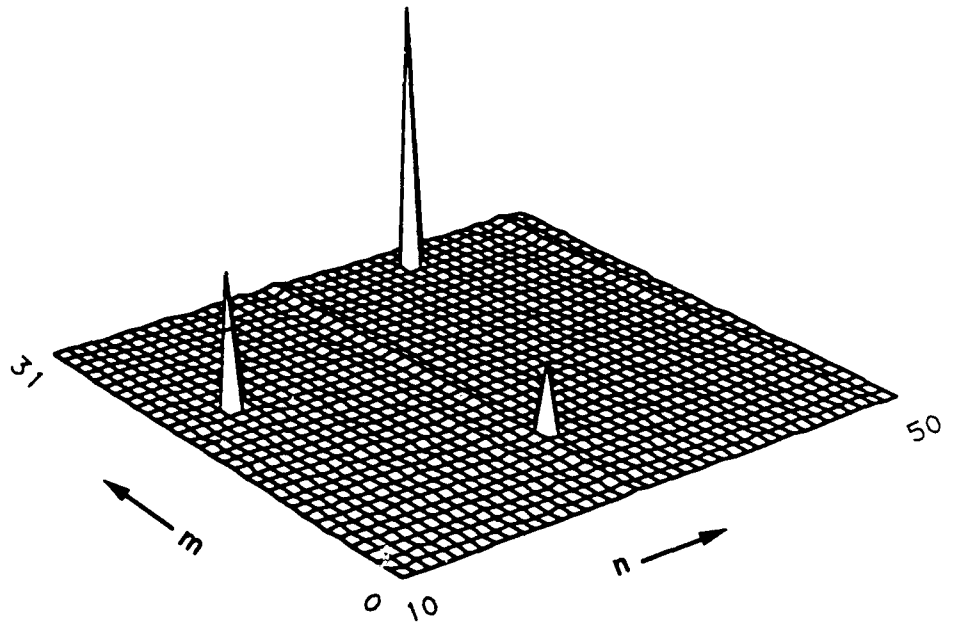

(b)

**Figure 4.** The real part of a noise–free signal with the Gabor representation using the Gaussian window function with three non-zero coefficients for (a) $\tau^2 = 1$ and (b) $\tau^2 = 0.5$. The locations and values of the coefficients are indicated in the figures. The sampling rate was $L = 64$ samples per second.

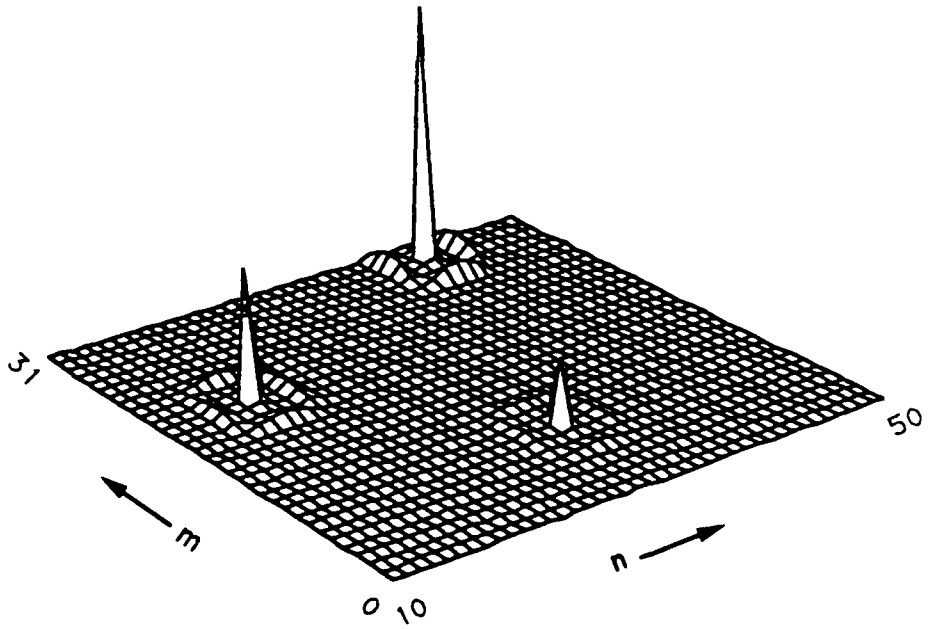
$r(m,n)$ decays exponentially with $m$ and $n$, reasonable values of $M$ and $N$ can be chosen to yield nearly optimal results. This procedure was implemented with $M = N = 5$.

The magnitudes of the coefficients computed by the two methods for the noise-free signals in Figure 4 are shown in Figure 5 for $\tau^2 = 1$ and in Figure 6 for $\tau^2 = 0.5$. Note that the sidelobes that result from filtering the sufficient statistics could be removed by using a larger filter. For example, when $\tau = 0.5$, the sidelobes in the time index disappear but the sidelobes in the frequency index appear. This is because the sufficient statistics are now more correlated in the frequency index as seen from (31). A rectangular filter which is longer in the frequency axis would solve this problem.

Complex zero-mean Gaussian white noise of variance $\sigma^2 = 2$ per sample ($\sigma^2 = 1$ for each the real part and the imaginary part, which are uncorrelated) was added to the signals in Figure 4 and the real part of these signals are shown in Figure 7. The resulting coefficients are depicted in Figure 8 and Figure 9. As seen in Figure 8, the results for the maximum likelihood method are only slightly better than those for the biorthogonal function method for $\tau^2 = 1$. However, the maximum likelihood estimates for $\tau^2 = 0.5$ shown in Figure 9 clearly have better noise properties than the other estimates. Heuristically, because of the mismatching involved in using the biorthogonal function, it seems reasonable that the maximum likelihood method would be better when the noise is large or when $\tau^2$ is small. When the noise is negligible and $\tau^2$ is moderate, say $\tau^2 \geq 1$, there may be little to gain by using the maximum likelihood method, which is slightly more complicated to implement.

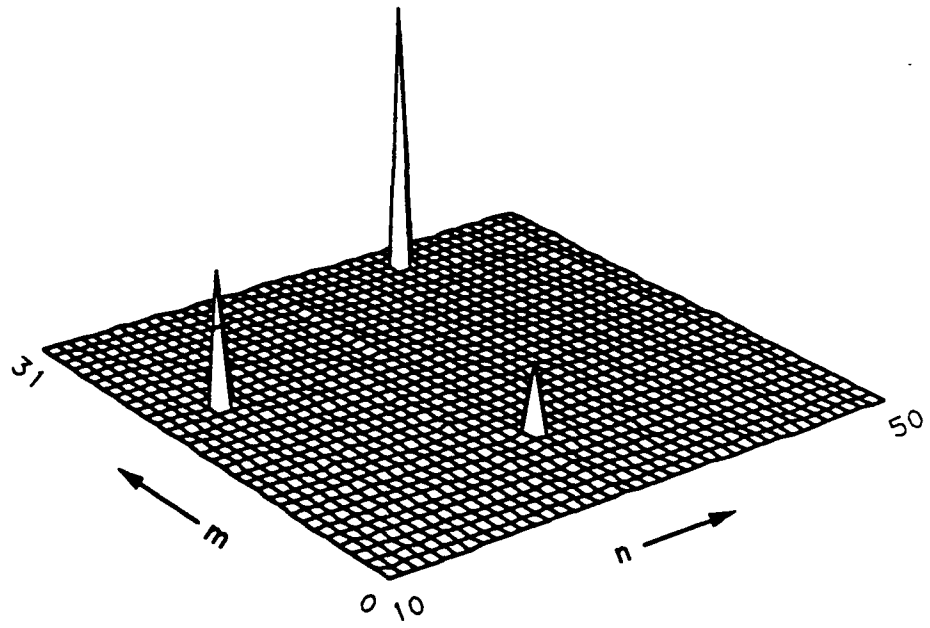



(a)

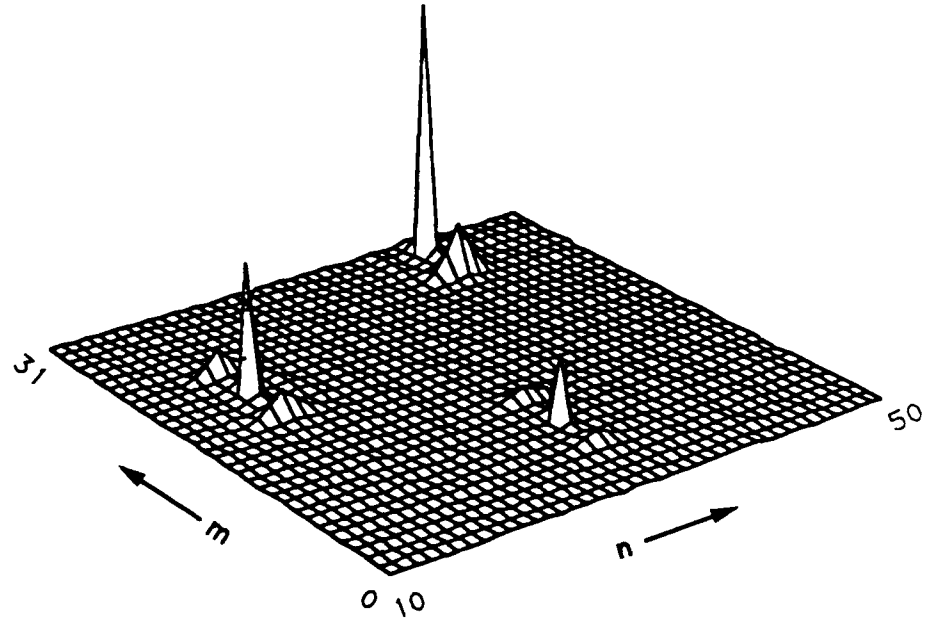



(b)

**Figure 5.** The magnitudes of the coefficients for the signal in Figure 4(a) using (a) the biorthogonal function method and (b) the maximum likelihood method.

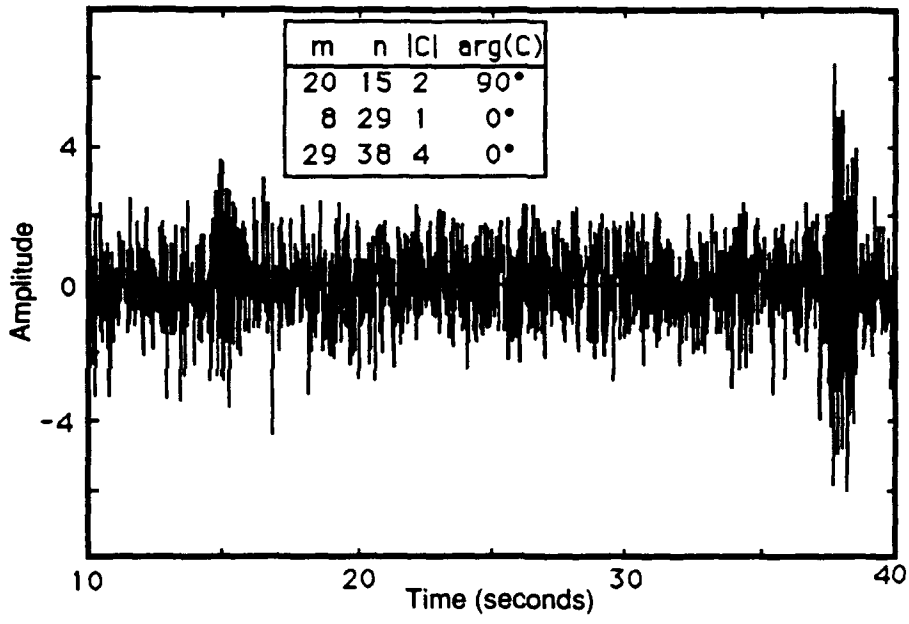



(a)

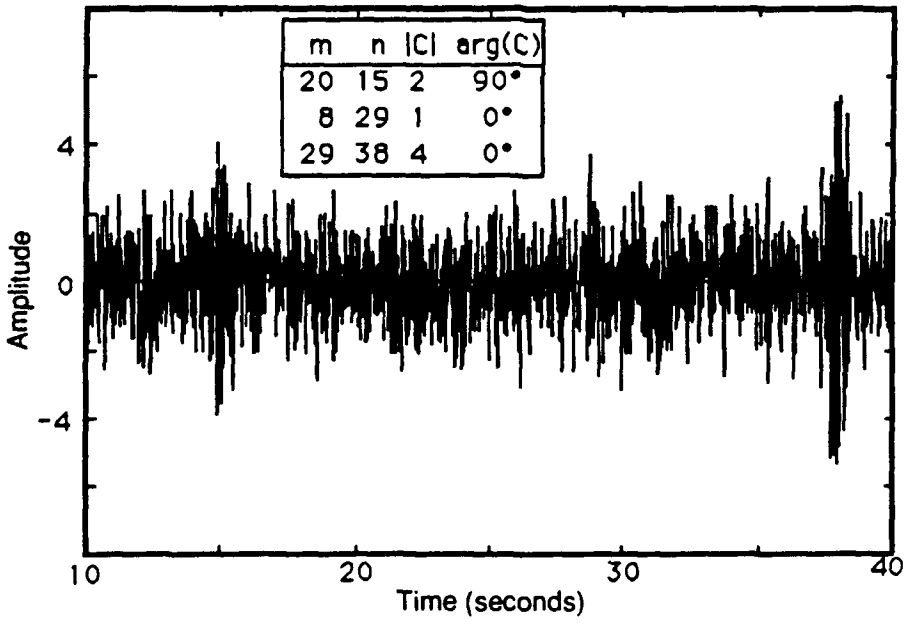



(b)

**Figure 6.** The magnitudes of the coefficients for the signal in Figure 4(b) using (a) the biorthogonal function method and (b) the maximum likelihood method.

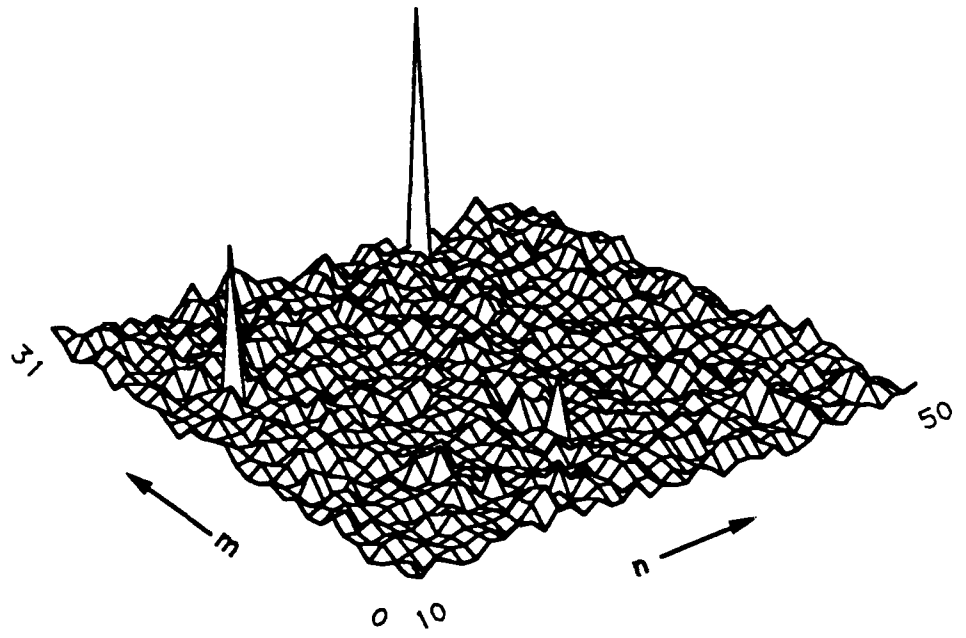



(a)

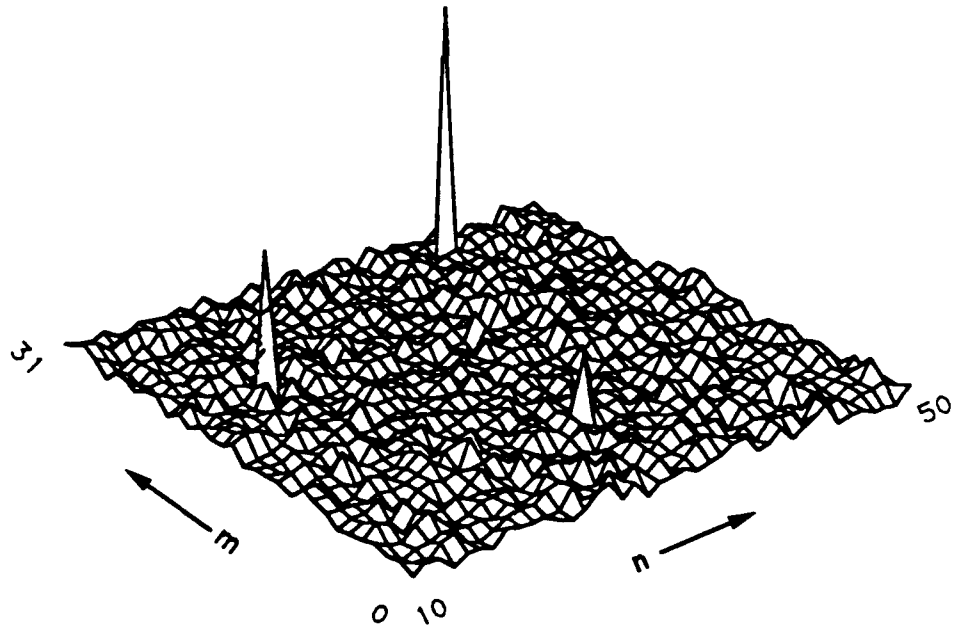



(b)

**Figure 7.** The real part of the signals in Figure 4 for (a) $\tau^2 = 1$ and (b) $\tau^2 = 0.5$ with additive complex white Gaussian noise with variance $\sigma^2 = 2$ per complex sample. The sampling rate was $L = 64$ samples per second.

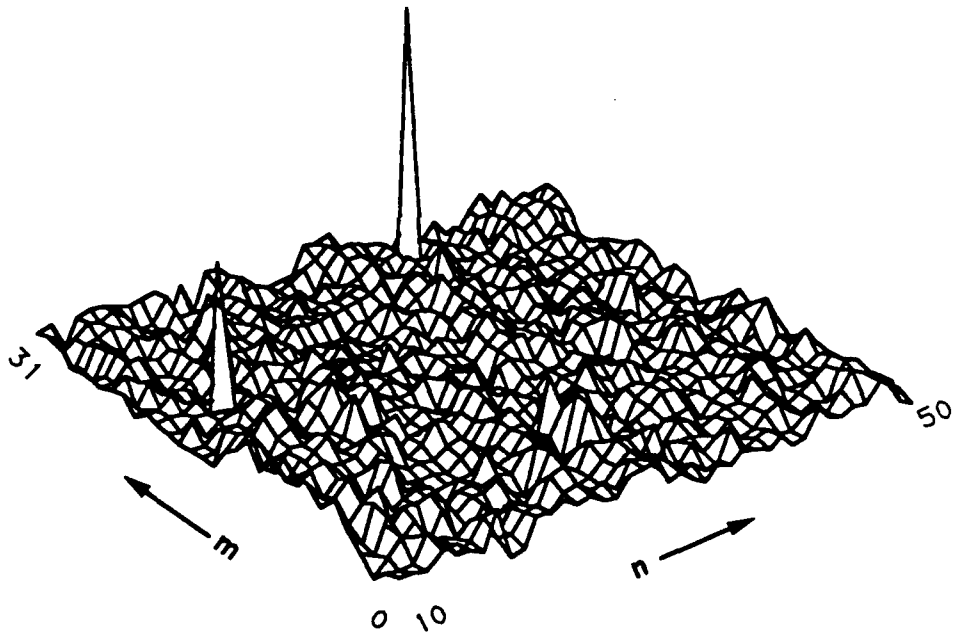



(a)

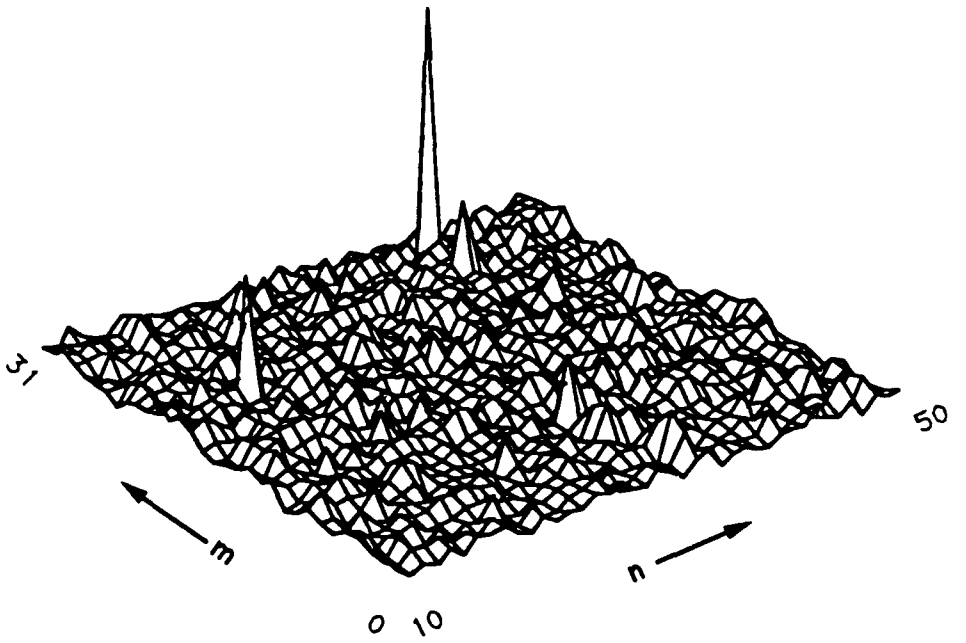



(b)

**Figure 8.** The magnitudes of the coefficients for the signal in Figure 7(a) using (a) the biorthogonal function method and (b) the maximum likelihood method.




(a)




(b)

**Figure 9.** The magnitudes of the coefficients for the signal in Figure 7(b) using (a) the biorthogonal function method and (b) the maximum likelihood method.

## 4.4 The Exponential Window Function

### 4.4.1 Introduction

While the Gaussian window function is well suited for signals which are made up of pulses with continuous and unimodal envelopes, it can not model well discontinuities in a signal, such as the type encountered in transient signal analysis. For this reason, the exponential window was investigated in [4]. Exponentially decaying sinusoids have been previously used to model transient signals, especially in relation to Prony's method [11,14,18]. The exponential window function $g_e(t)$ is given by

$$g(t) = \sqrt{2\lambda}\exp(-\lambda t)u(t), \tag{33}$$

where $\lambda$ is a parameter and $u(t)$ is the unit step function. The leading constant in (33) is chosen so that $\int |g(t)|^2 dt = 1$.

### 4.4.2 The Biorthogonal Function Method

The biorthogonal function for the exponential window is [3,4]

$$\gamma_e(t) = \frac{\exp(\lambda t)}{\sqrt{2\lambda}}[-u(t+1) + 2u(t) - u(t-1)].$$

This function is zero outside the interval $[-1 : 1]$ which simplifies the computation of the Gabor coefficients:

$$C_{mn} = \int_{-1}^{1} s(t+n)\gamma_e(t)\exp(-j2\pi mt)dt. \tag{34}$$

The window function and the biorthogonal function for $\lambda = 0.5$ are shown in Figure 10, and those for $\lambda = 4$ are shown in Figure 11.

A further simplification is [3,4]

$$C_{mn} = D_{mn} - \exp(-\lambda)D_{m,n-1}$$

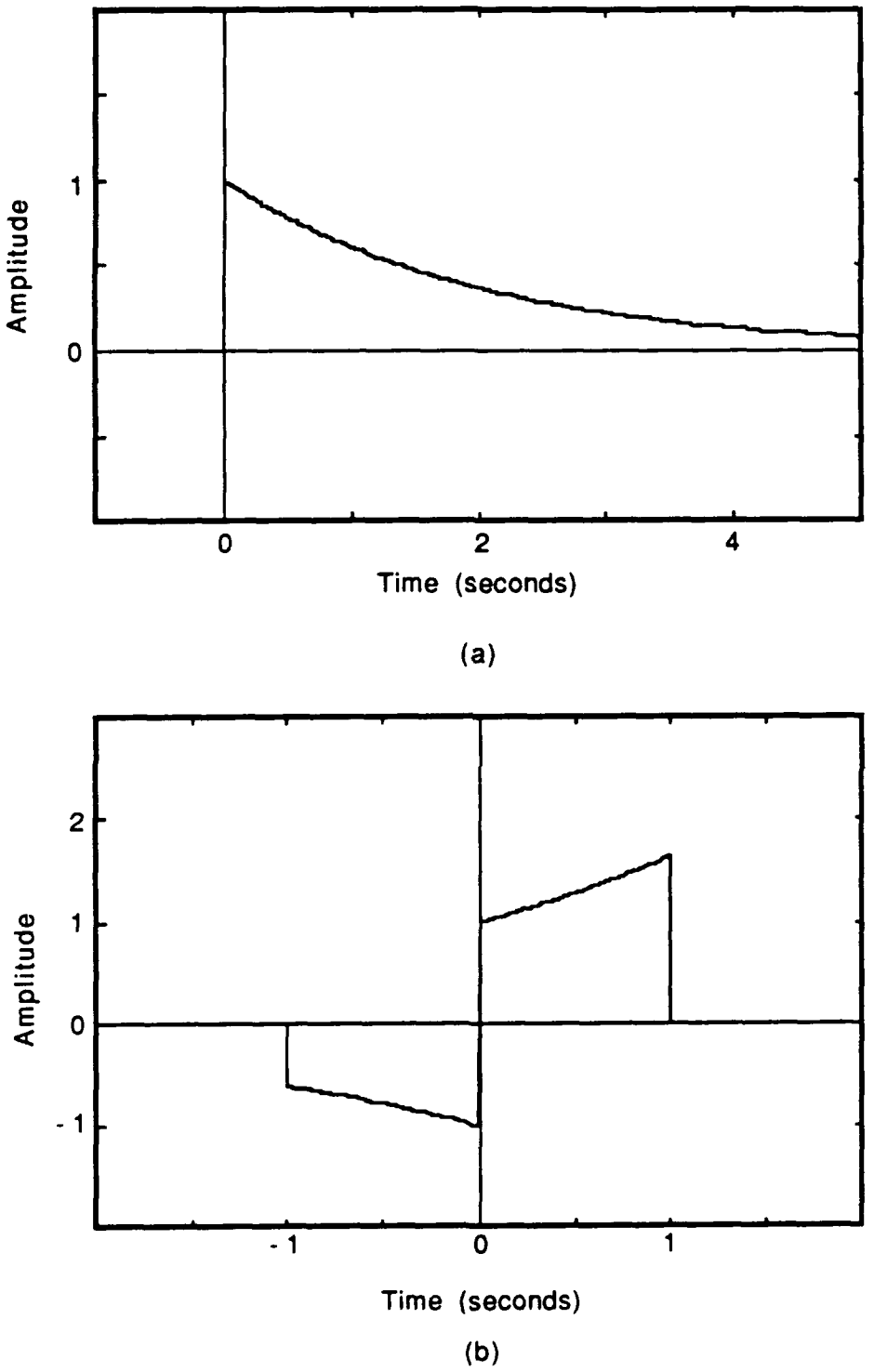
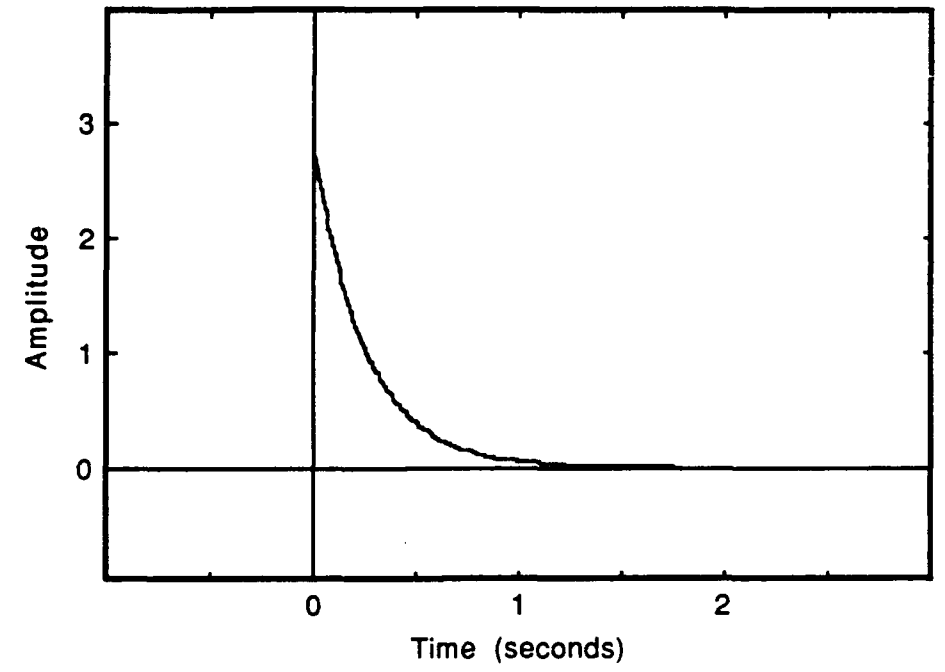
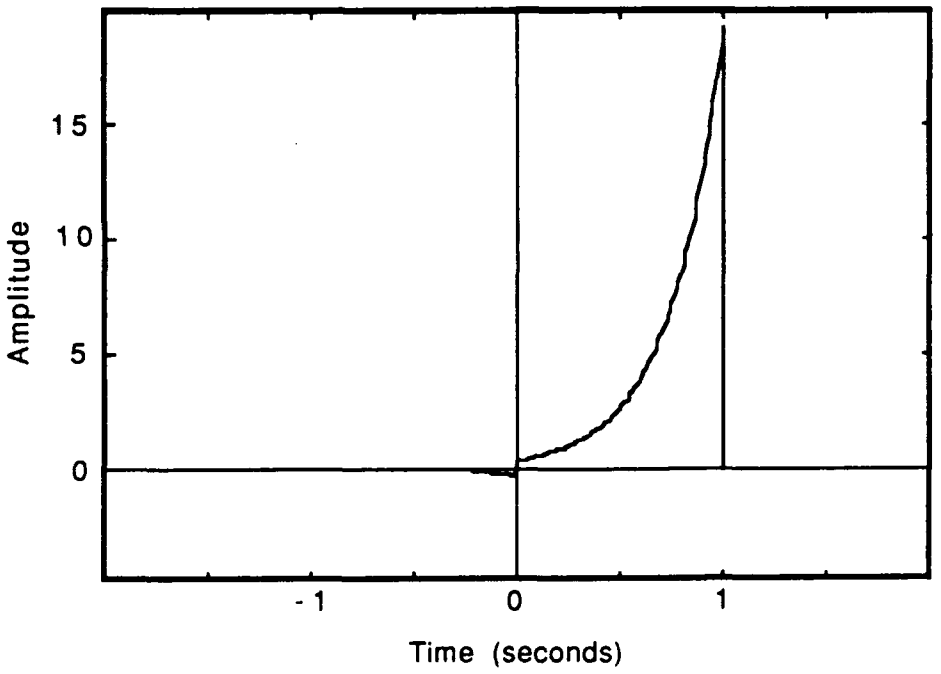



**Figure 10.** (a) The exponential window function $g_e(t)$ and (b) the biorthogonal function $\gamma_e(t)$ for $\lambda = 0.5$.




(a)




(b)

**Figure 11.** (a) The exponential window function $g_e(t)$ and (b) the biorthogonal function $\gamma_e(t)$ for $\lambda = 4$.

where

$$D_{mn} = \frac{1}{\sqrt{2\lambda}} \int_0^1 s(t+n)\exp(\lambda t)\exp(-j2\pi mt)dt.$$

In [4], the statistics of the Gabor coefficients in real white Gaussian noise with autocorrelation function $N_0\delta(t-s)$ are derived and the results are included here for later reference. Suppose the $NM$ coefficients $\{C_{mn}; 0 \le m \le M-1, 0 \le n \le N-1\}$ are desired, and let $C^R_{mn} = \Re\{C_{mn}\}$ and $C^I_{mn} = \Im\{C_{mn}\}$. Let $\mathbf{C}_n$ be the $(2M-1)$ component vector

$$\mathbf{C}_n = [C^R_{0,n}\ C^R_{1,n} \cdots C^R_{M-1,n}\ C^I_{1,n} \cdots C^I_{M-1,n}]^T \tag{35}$$

and let $\mathbf{C}$ be the $(2M-1)N$ component vector

$$\mathbf{C} = [\mathbf{C}_0^T\ \mathbf{C}_1^T \cdots \mathbf{C}_{N-1}^T]^T. \tag{36}$$

Since $C_{0,n}$ is real, $C^I_{0,n}$ is omitted in (35). The covariance matrix for $\mathbf{C}$ is

$$E[\mathbf{C}\mathbf{C}^T] = \frac{N_0[\exp(2\lambda) - \exp(-2\lambda)]}{2\lambda}\mathbf{P} \otimes \mathbf{Q}, \tag{37}$$

where $= \otimes$ denotes the Kronecker product, the $N \times N$ matrix $\mathbf{P}$ is the tridiagonal Toeplitz matrix

$$\mathbf{P} = \begin{bmatrix} 1 & -\alpha & & 0 \\ -\alpha & \ddots & \ddots & \\ & \ddots & \ddots & -\alpha \\ 0 & & -\alpha & 1 \end{bmatrix} \tag{38}$$

with

$$\alpha = [\exp(\lambda) + \exp(-\lambda)]^{-1}, \tag{39}$$

and the $(2M-1) \times (2M-1)$ matrix $\mathbf{Q}$ is given by

$$\begin{aligned}
(\mathbf{Q})_{mk} &= \frac{\lambda}{4\lambda^2 + 4\pi^2(m-k)^2} + \frac{\lambda}{4\lambda^2 + 4\pi^2(m+k)^2}, && 0 \le m,k \le M-1 \\
(\mathbf{Q})_{M-1+m,k} &= \frac{\pi(m+k)}{4\lambda^2 + 4\pi^2(m+k)^2} + \frac{\pi(m-k)}{4\lambda^2 + 4\pi^2(m-k)^2}, && 0 \le k \le M-1 \\
& && 1 \le m \le M-1 \\
(\mathbf{Q})_{k,M-1+m} &= \frac{\pi(m+k)}{4\lambda^2 + 4\pi^2(m+k)^2} + \frac{\pi(m-k)}{4\lambda^2 + 4\pi^2(m-k)^2}, && 0 \le k \le M-1 \\
& && 1 \le m \le M-1 \\
(\mathbf{Q})_{M-1+m,M-1+k} &= \frac{\lambda}{4\lambda^2 + 4\pi^2(m-k)^2} - \frac{\lambda}{4\lambda^2 + 4\pi^2(m+k)^2}, && 1 \le k,m \le M-1.
\end{aligned} \tag{40}$$

### 4.4.3 The Maximum Likelihood Method for Finite Observation Intervals

Consider a finite observation interval $I = [0, N]$ and let the allowable frequency indices $m$ satisfy $0 \leq m \leq M-1$. The received signal $r(t)$ consists of a real signal $s(t)$ given by

$$s(t) = \sum_{n=0}^{N-1} A_{0,n} g_e(t-n) + \sum_{m=1}^{M-1} \sum_{n=0}^{N-1} A_{mn} g_e(t-n) \cos(2\pi mt + \theta_{mn}) \tag{41}$$

corrupted by additive real, white Gaussian noise $n(t)$ with $E[n(t)n(t+\tau)] = N_0\delta(\tau)$. Below, the method described briefly in Section 2.6 for handling the real signal case is developed in detail using the example of the exponential window function. A different derivation is possible which uses complex sufficient statistics and matrices, but the approach taken here is to use all real signal processing.

The signal in (41) can be rewritten in the form

$$s(t) = \sum_{n=0}^{N-1} U_{0,n} g_e(t-n) + \sum_{m=1}^{M-1} \sum_{n=0}^{N-1} [U_{mn} g_e(t-n) \cos(2\pi mt) + V_{mn} g(t-n) \sin(2\pi mt)] \tag{42}$$

where

$$\begin{aligned} U_{0,n} &= A_{0,n} \\ U_{mn} &= A_{mn} \cos(\theta_{mn}) \\ V_{mn} &= A_{mn} \sin(\theta_{mn}). \end{aligned} \tag{43}$$

For a given set of coefficients $\{U_{mn}, V_{mn}\}$, the likelihood of the received waveform is

$$P[\{r_t, t \in I\} | \{U_{mn}, V_{mn}\}] = \kappa \exp[\Omega(\{U_{mn}, V_{mn}\})/N_0] \tag{44}$$

where

$$\Omega(\{U_{mn}, V_{mn}\}) = 2\int_I s_t(\{U_{mn}, V_{mn}\}) dr_t - \int_I |s_t(\{U_{mn}, V_{mn}\})|^2 dt \tag{45}$$

and $\kappa$ is a positive constant that does not depend on $\{U_{mn}, V_{mn}\}$. Set

$$\begin{aligned} I_{mn} &= \int_I g_e(t-n) \cos(2\pi mt) dr_t \\ Q_{mn} &= \int_I g_e(t-n) \sin(2\pi mt) dr_t. \end{aligned} \tag{46}$$

Then

$$\int_I s_t(\{U_{mn}, V_{mn}\})dr_t = \sum_{n=0}^{N-1} U_{0,n}I_{0,n} + \sum_{m=1}^{M-1}\sum_{n=0}^{N-1}[U_{mn}I_{mn} + V_{mn}Q_{mn}]. \tag{47}$$

It follows from (44)–(47) that $\{I_{mn}; 0 \le m \le N-1, 0 \le n \le N-1\}$ and $\{Q_{mn}; 1 \le m \le N-1, 0 \le n \le N-1\}$ are jointly sufficient statistics for the distribution (44). Define the following signal cross-correlations:

$$\begin{aligned}
r^{II}[(k,l),(n,m)] &= \int_I g_e(t-l)\cos(2\pi kt)g_e(t-n)\cos(2\pi mt)dt, && \begin{matrix}0 \le k,m \le M-1\\ 0 \le l,n \le N-1\end{matrix}\\
r^{IQ}[(k,l),(n,m)] &= \int_I g_e(t-l)\cos(2\pi kt)g_e(t-n)\sin(2\pi mt)dt, && \begin{matrix}0 \le k \le M-1\\ 1 \le m \le M-1\\ 0 \le l,n \le N-1\end{matrix}\\
r^{QI}[(k,l),(n,m)] &= \int_I g_e(t-l)\sin(2\pi kt)g_e(t-n)\cos(2\pi mt)dt, && \begin{matrix}1 \le k \le M-1\\ 0 \le m \le M-1\\ 0 \le l,n \le N-1\end{matrix}\\
r^{QQ}[(k,l),(n,m)] &= \int_I g_e(t-l)\sin(2\pi kt)g_e(t-n)\sin(2\pi mt)dt, && \begin{matrix}1 \le k,m \le M-1\\ 0 \le l,n \le N-1\end{matrix}
\end{aligned} \tag{48}$$

Now, the second term on the right hand side of (45) can be written as

$$\begin{aligned}
\int_I |s_t(\{U_{mn}, V_{mn}\})|^2 dt = &\sum_{k=0}^{M-1}\sum_{l=0}^{N-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} U_{kl}r^{II}[(k,l),(n,m)]U_{mn}\\
&+2\sum_{k=1}^{M-1}\sum_{l=0}^{N-1}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} V_{kl}r^{QI}[(k,l),(n,m)]U_{mn}\\
&+\sum_{k=1}^{M-1}\sum_{l=0}^{N-1}\sum_{m=1}^{M-1}\sum_{n=0}^{N-1} V_{kl}r^{QQ}[(k,l),(n,m)]V_{mn}.
\end{aligned} \tag{49}$$

Define the $(2M-1)$ component real vector $\mathbf{W}_n$ by

$$\mathbf{W}_n = \left[U_{0,n}\ U_{1,n}\cdots U_{M-1,n}\ V_{1,n}\cdots V_{M-1,n}\right]^T \tag{50}$$

and the $(2M-1)N$ component real vector $\mathbf{W}$ by

$$\mathbf{W} = \left[\mathbf{W}_0^T\ \mathbf{W}_1^T\cdots\mathbf{W}_{N-1}^T\right]^T. \tag{51}$$

Similarly, define the $(2M-1)$ component vector $\mathbf{T}_n$ of sufficient statistics by

$$\mathbf{T}_n = \left[I_{0,n}\ I_{1,n}\cdots I_{M-1,n}\ Q_{1,n}\cdots Q_{M-1,n}\right]^T \tag{52}$$

and the $(2M-1)N$ component vector $\mathbf{T}$ by

$$\mathbf{T} = \left[\mathbf{T}_0^T \; \mathbf{T}_1^T \cdots \mathbf{T}_{N-1}^T\right]^T. \tag{53}$$

Arrange the $(2M-1)N \times (2M-1)N$ signal cross-correlation matrix $\mathbf{R}$ according to

$$\begin{aligned}
(\mathbf{R})_{l(2M-1)+k+1,n(2M-1)+m+1} &= r^{II}[(k,l),(n,m)], && \substack{0\le k,m\le M-1\\ 0\le l,n\le N-1}\\
(\mathbf{R})_{l(2M-1)+M+k,n(2M-1)+m+1} &= r^{QI}[(k,l),(m,n)], && \substack{1\le k\le M-1 \quad 0\le m\le M-1\\ 0\le l,n\le N-1}\\
(\mathbf{R})_{l(2M-1)+k+1,n(2M-1)+M+m} &= r^{IQ}[(k,l),(n,m)], && \substack{1\le k\le M-1 \quad 0\le m\le M-1\\ 0\le l,n\le N-1}\\
(\mathbf{R})_{l(2M-1)+M+k,n(2M-1)+M+m} &= r^{QQ}[(k,l),(n,m)], && \substack{1\le k,m\le M-1\\ 0\le l,n\le N-1}.
\end{aligned} \tag{54}$$

Now, (44) can be expressed as

$$\Omega(\{U_{mn}, V_{mn}\}) = 2\mathbf{W}^T\mathbf{T} - \mathbf{W}^T\mathbf{R}\mathbf{W}. \tag{55}$$

The maximum likelihood estimate of $\mathbf{W}$, $\widehat{\mathbf{W}}_{\mathrm{MLE}}$, maximizes (55) and is the solution of the set of linear equations

$$\mathbf{R}\mathbf{W} = \mathbf{T},$$

which, for $\mathbf{R}$ invertible, has solution

$$\widehat{\mathbf{W}}_{\mathrm{MLE}} = \mathbf{R}^{-1}\mathbf{T}. \tag{56}$$

It is straightforward to show that

$$\begin{aligned}
E\left[\widehat{\mathbf{W}}_{\mathrm{MLE}}\right] &= \mathbf{W}\\
E\left[\left(\widehat{\mathbf{W}}_{\mathrm{MLE}} - \mathbf{W}\right)\left(\widehat{\mathbf{W}}_{\mathrm{MLE}} - \mathbf{W}\right)^T\right] &= N_0\mathbf{R}^{-1}.
\end{aligned} \tag{57}$$

The procedure outlined above is valid for any window function. When the exponential window function is used with the observation interval $I = [0, N]$, the signal cross-correlations

in (48) are as follows:

$$r^{II}[(k,l),(m,n)] = 2\lambda^2 r_1(l,n)\left[\frac{1}{4\lambda^2+4\pi^2(m-k)^2}+\frac{1}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{IQ}[(k,l),(m,n)] = \lambda r_1(l,n)\left[\frac{2\pi(m-k)}{4\lambda^2+4\pi^2(m-k)^2}+\frac{2\pi(k+m)}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{QI}[(k,l),(m,n)] = \lambda r_1(l,n)\left[\frac{2\pi(k-m)}{4\lambda^2+4\pi^2(m-k)^2}+\frac{2\pi(k+m)}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{QQ}[(k,l),(m,n)] = 2\lambda^2 r_1(l,n)\left[\frac{1}{4\lambda^2+4\pi^2(m-k)^2}-\frac{1}{4\lambda^2+4\pi^2(m+k)^2}\right],$$

where $r_1(l,n)$ is given by

$$r_1(l,n) = e^{\lambda(l+n)}\left[e^{-2\lambda(l\vee n)}-e^{-2\lambda N}\right]$$

with $x \vee y = \max(x,y)$. Because the signal cross-correlations factor into a product of two terms, one which depends only on the delays $l$ and $n$ and one which depends only on the frequencies $k$ and $m$, the signal cross-correlation matrix $\mathbf{R}$ factors into the Kronecker product

$$\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2 \tag{58}$$

where $\mathbf{R}_1$ is the $N \times N$ matrix with entries

$$(\mathbf{R}_1)_{l,n} = \exp\{\lambda(l+n)\}\left[\exp\{-2\lambda(l\vee n)\}-\exp\{-2\lambda N\}\right]$$

and where $\mathbf{R}_2$ is the $(2M-1)\times(2M-1)$ matrix with entries

$$\begin{aligned}
(\mathbf{R}_2)_{mk} &= \frac{2\lambda^2}{4\lambda^2+4\pi^2(m-k)^2}+\frac{2\lambda^2}{4\lambda^2+4\pi^2(m+k)^2} && 0\le k,m\le M-1\\
(\mathbf{R}_2)_{k,M-1+m} &= \frac{2\lambda\pi(m+k)}{4\lambda^2+4\pi^2(m+k)^2}+\frac{2\lambda\pi(m-k)}{4\lambda^2+4\pi^2(m-k)^2} && \begin{matrix}0\le k\le M-1\\ 1\le m\le M-1\end{matrix}\\
(\mathbf{R}_2)_{M-1+m,k} &= \frac{2\lambda\pi(m+k)}{4\lambda^2+4\pi^2(m+k)^2}+\frac{2\lambda\pi(m-k)}{4\lambda^2+4\pi^2(m-k)^2} && \begin{matrix}0\le k\le M-1\\ 1\le m\le M-1\end{matrix}\\
(\mathbf{R}_2)_{M-1+m,M-1+k} &= \frac{2\lambda^2}{4\lambda^2+4\pi^2(m-k)^2}-\frac{2\lambda^2}{4\lambda^2+4\pi^2(m+k)^2} && 1\le k,m\le M-1.
\end{aligned} \tag{59}$$

Note that the matrix $\mathbf{R}_2$ and the matrix $\mathbf{Q}$ in (40) are identical to within a scale factor. However, $\mathbf{R}_2$ is part of the covariance matrix of the sufficient statistics, or, through (57),

related to the inverse covariance matrix of the maximum likelihood coefficient estimates. On the other hand, **Q** is part of the covariance matrix of the coefficients estimated by the biorthogonal function.

The fact that signal cross correlation matrix factors into a Kronecker product simplifies the processing to compute $\widehat{\mathbf{W}}_{\mathrm{MLE}} = \mathbf{R}^{-1}\mathbf{T}$ since

$$\mathbf{R}^{-1} = (\mathbf{R}_1 \otimes \mathbf{R}_2)^{-1} = \mathbf{R}_1^{-1} \otimes \mathbf{R}_2^{-1}. \tag{60}$$

### 4.4.4 Examples using the Exponential Window

A noise-free signal with three nonzero coefficients using the exponential window function for $\lambda = 0.5$ is shown in Figure 12 along with the same signal corrupted by additive zero-mean white Gaussian noise with a variance of $\sigma^2 = 1$ per sample. The magnitudes of the Gabor coefficients computed for each of the two methods for $M = N = 16$ for the noise-free case and the noisy case are shown in Figure 13 and Figure 14. The sampling rate was $L = 1024$ samples per second. Figure 15 shows a signal and its noise corrupted version with same coefficients as that in Figure 12 but with $\lambda = 4$. Figure 16 and Figure 17 show the corresponding coefficients. Overall, the quality of the noise free estimates is nearly identical. For this example, the overall quality of the estimates in the noisy case is roughly the same for $\lambda = 0.5$ but the maximum likelihood method is clearly superior for $\lambda = 4$. This agrees with the intuition that the biorthogonal function method performs poorly for large values of $\lambda$ in the presence of noise because it amplifies the portion of the signal where the window function is smallest, thus amplifying the noise.

In order to quantitatively compare the performance of the two schemes, the variance of the estimates of $U_{mn}$ and $V_{mn}$ in (43) are compared. From (37)-(40), the estimation

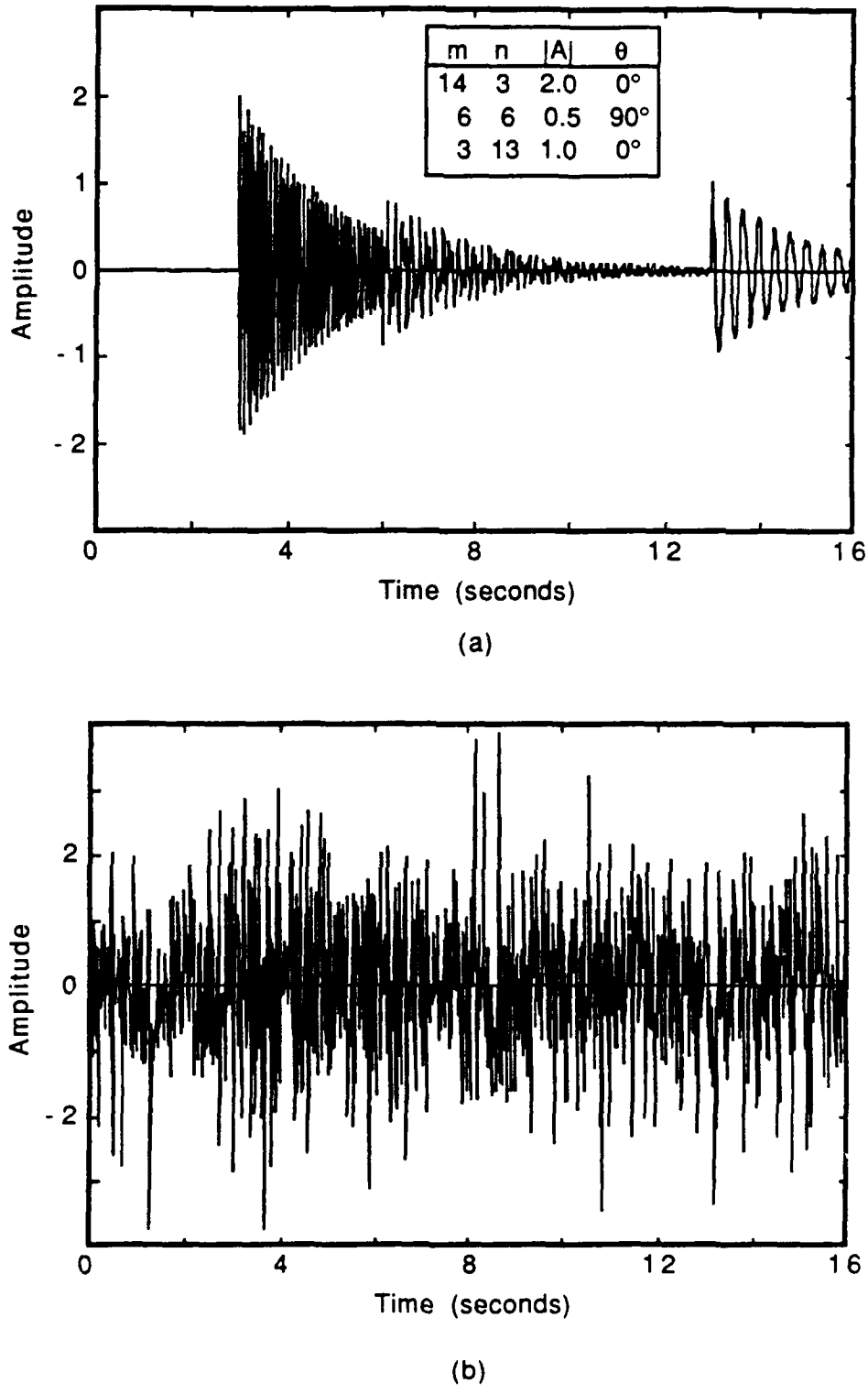
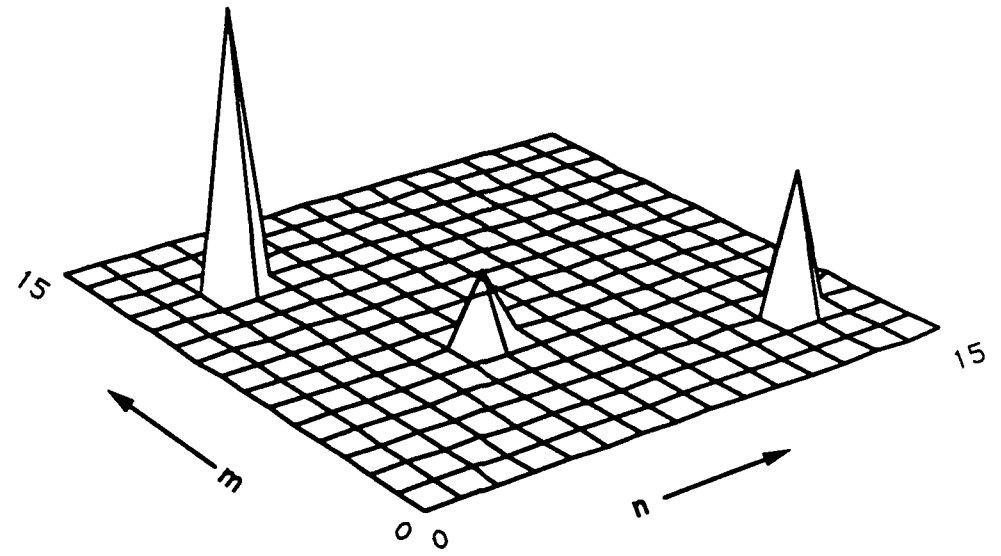
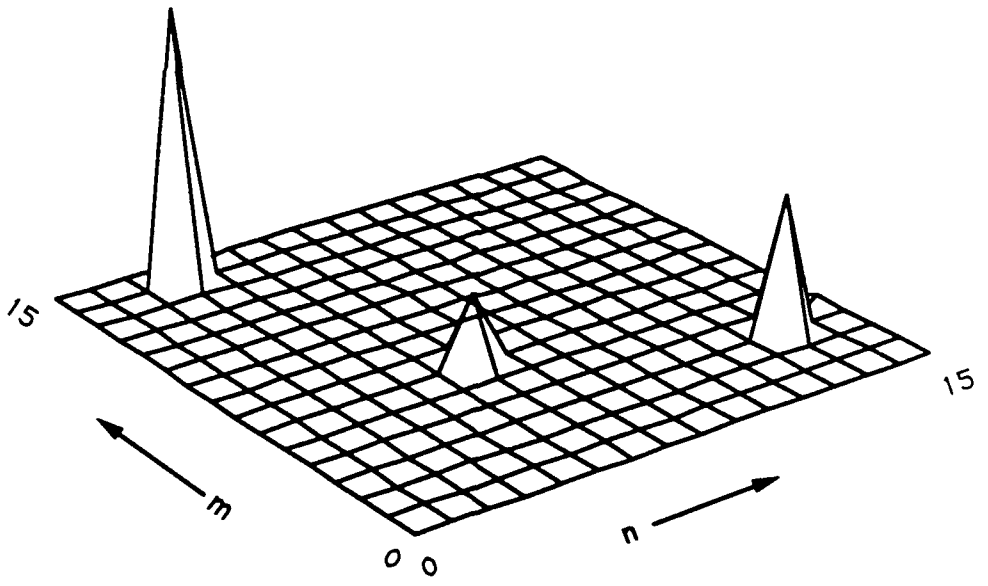



**Figure 12.** A real signal with the representation in (41) using the exponential window function with $\lambda = 0.5$ (a) without added noise and (b) with white Gaussian noise with variance $\sigma^2 = 1$ added per sample. The three non-zero coefficients are shown in (a). The sampling rate was $L = 1024$ samples per second for the processing but $L = 64$ samples per second was used for this figure.

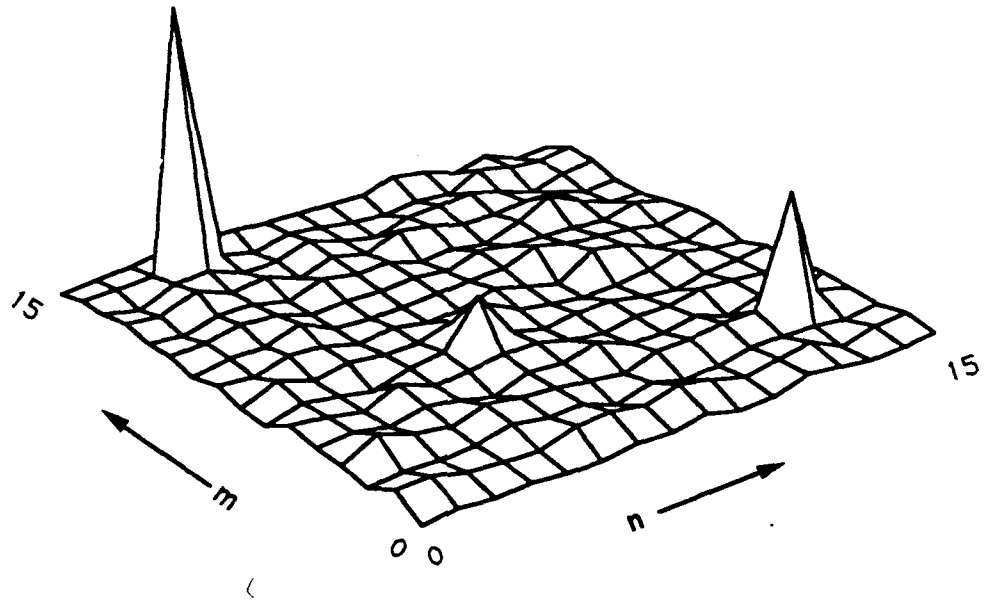



(a)

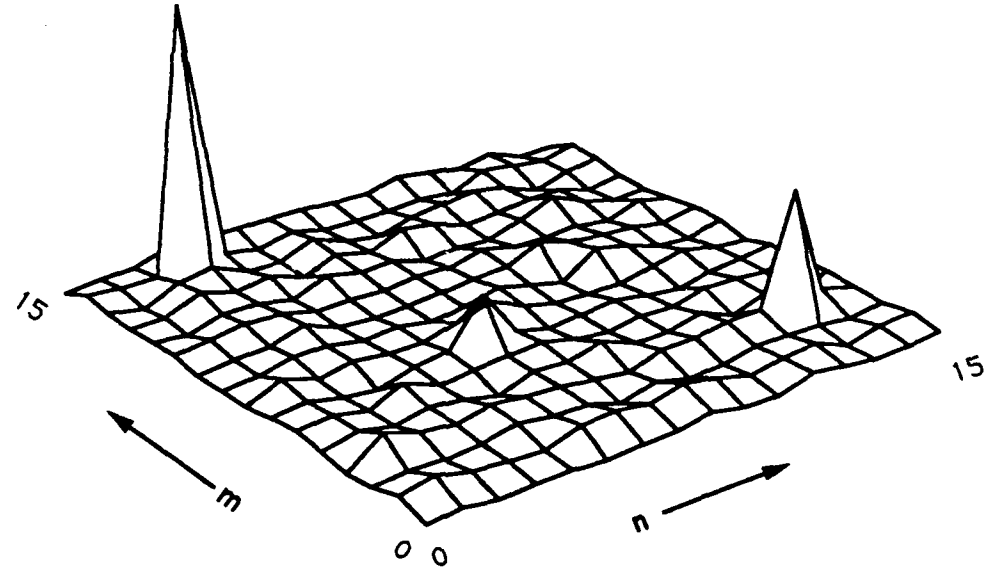



(b)

**Figure 13.** The magnitudes of the coefficients of the signal in Figure 12(a) computed using (a) the biorthogonal function method and (b) the maximum likelihood method. The sampling rate was $L = 1024$ samples per second.

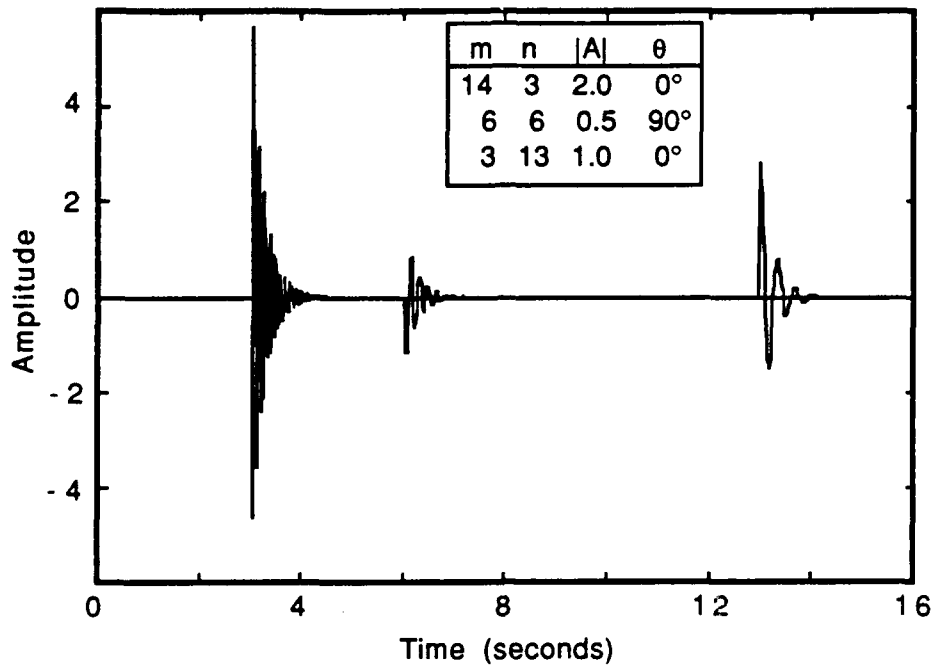



(a)




(b)

**Figure 14.** The magnitudes of the coefficients of the signal in Figure 12(b) computed using (a) the biorthogonal function method and (b) the maximum likelihood method. The sampling rate was $L = 1024$ samples per second.



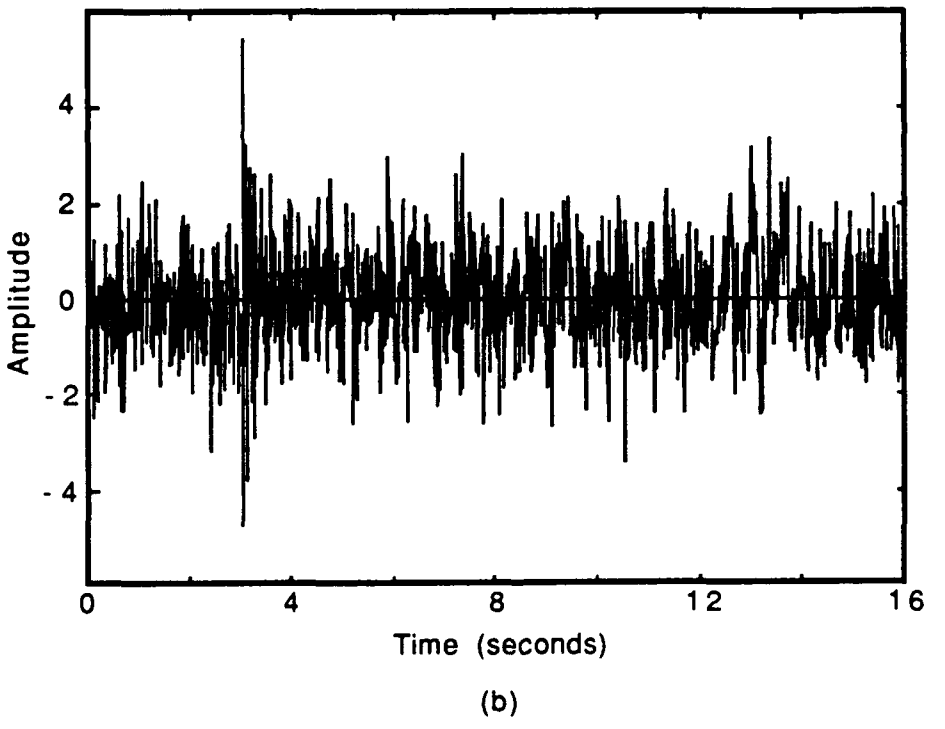
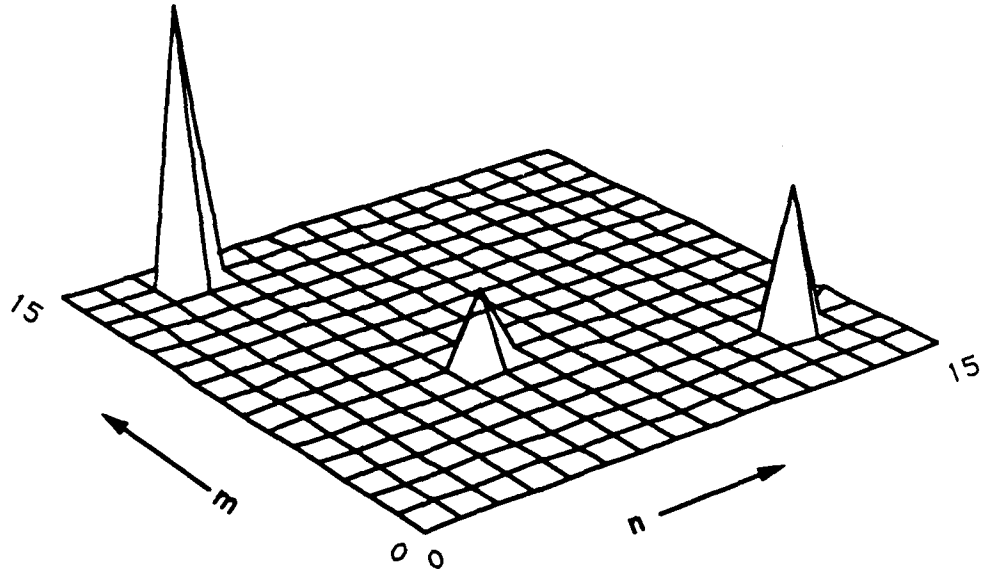
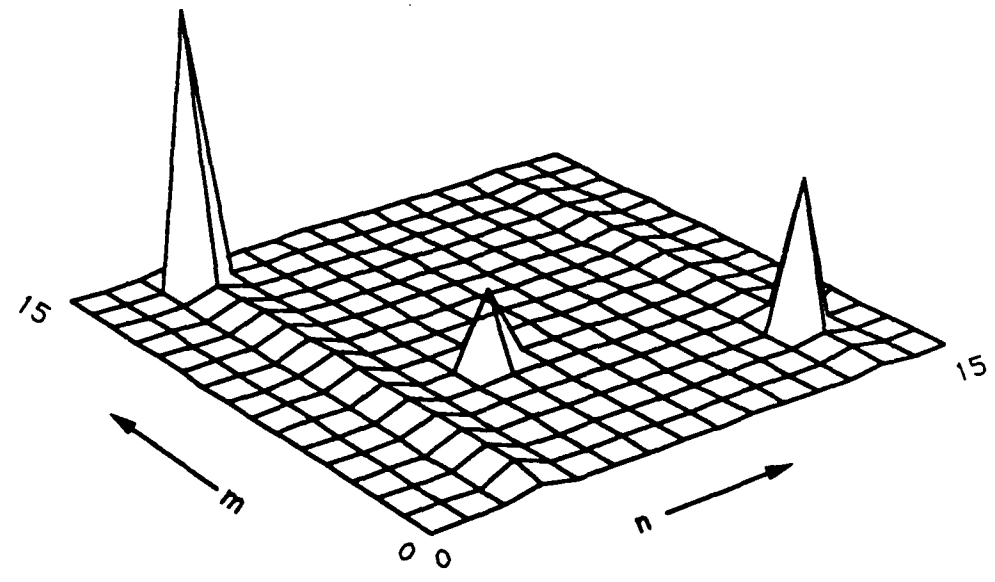




**Figure 15.** A real signal with the representation in (41) using the exponential window function with $\lambda = 4$ (a) without added noise and (b) with white Gaussian noise with variance $\sigma^2 = 1$ added per sample. The three non-zero coefficients are shown in (a). The sampling rate was $L = 1024$ samples per second for the processing but $L = 64$ samples per second was used for this figure.

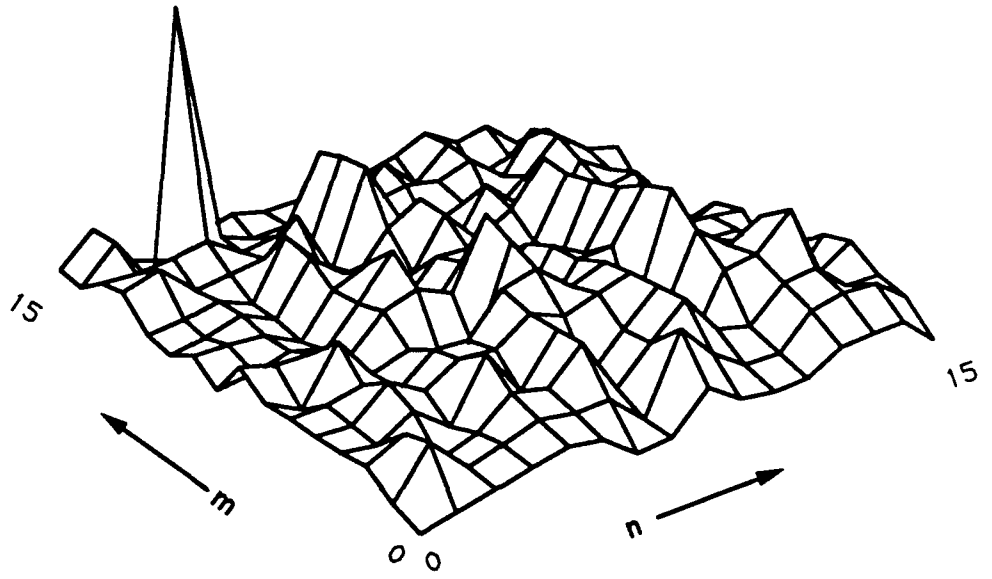



(a)

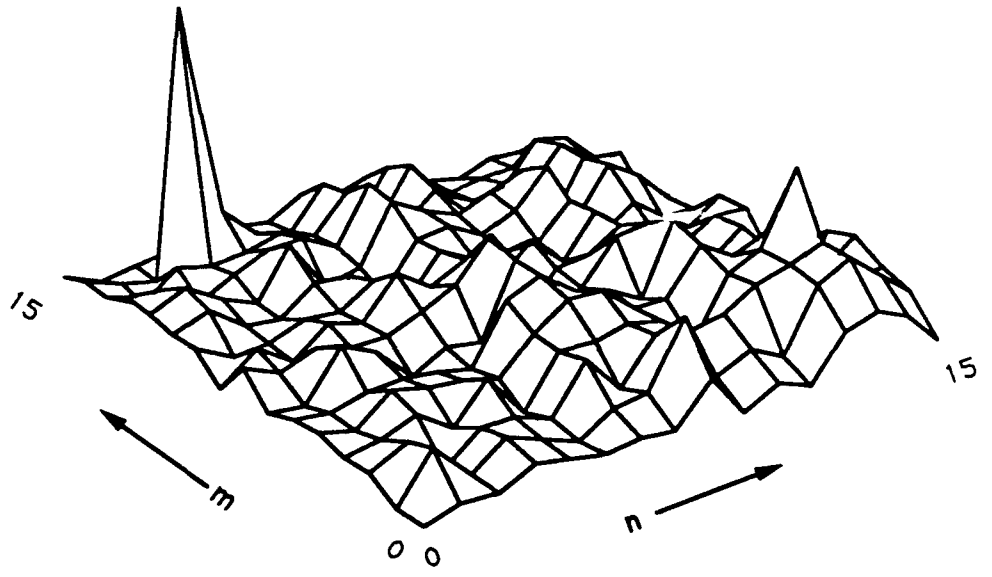



(b)

**Figure 16.** The magnitudes of the coefficients of the signal in Figure 15(a) computed using (a) the biorthogonal function method and (b) the maximum likelihood method. The sampling rate was $L = 1024$ samples per second.




(a)




(b)

**Figure 17.** The magnitudes of the coefficients of the signal in Figure 15(b) computed using (a) the biorthogonal function method and (b) the maximum likelihood method. The sampling rate was $L = 1024$ samples per second.

variance for the biorthogonal function method is

$$\begin{aligned}\mathrm{Var}[U_{mn}] &= \frac{N_0[\exp(2\lambda)-\exp(-2\lambda)]}{2\lambda}\left[\frac{1}{4\lambda}+\frac{\lambda}{4\lambda^2+16\pi^2m^2}\right]\\ \mathrm{Var}[V_{mn}] &= \frac{N_0[\exp(2\lambda)-\exp(-2\lambda)]}{2\lambda}\left[\frac{1}{4\lambda}-\frac{\lambda}{4\lambda^2+16\pi^2m^2}\right].\end{aligned} \tag{61}$$

The variances for the maximum likelihood estimates of $U_{mn}$ and $V_{mn}$ are found according to (57) to be the product of $N_0$ and the corresponding element in $\mathbf{R}^{-1}$ on the main diagonal. The inverse was computed for $M = N = 16$ using the simplified structure in (60). Let $\rho_U(m,n)$ and $\rho_V(m,n)$ be the reduction in estimation variance for $U_{mn}$ and $V_{mn}$ obtained using the maximum likelihood estimates instead of those obtained by the BFM, that is

$$\begin{aligned}\rho_U(m,n) &= \frac{\mathrm{Var}\left[(\hat{U}_{mn})_{\mathrm{BFM}}\right]}{\mathrm{Var}\left[(\hat{U}_{mn})_{\mathrm{MLE}}\right]}\\ \rho_V(m,n) &= \frac{\mathrm{Var}\left[(\hat{V}_{mn})_{\mathrm{BFM}}\right]}{\mathrm{Var}\left[(\hat{V}_{mn})_{\mathrm{MLE}}\right]}.\end{aligned}$$

Define the following:

$$\begin{aligned}\rho_{\max} &= \max\left\{\max_{\substack{0\le m\le M-1\\ 0\le n\le N-1}}\rho_U(m,n),\ \max_{\substack{1\le m\le M-1\\ 0\le n\le N-1}}\rho_V(m,n)\right\}\\ \rho_{\min} &= \min\left\{\min_{\substack{0\le m\le M-1\\ 0\le n\le N-1}}\rho_U(m,n),\ \min_{\substack{1\le m\le M-1\\ 0\le n\le N-1}}\rho_V(m,n)\right\}\\ \rho_{\mathrm{avg}} &= \frac{1}{(2M-1)N}\left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\rho_U(m,n)+\sum_{m=1}^{M-1}\sum_{n=0}^{N-1}\rho_V(m,n)\right]\\ \rho_{\mathrm{dB-avg}} &= \frac{1}{(2M-1)N}\left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}10\log_{10}\rho_U(m,n)+\sum_{m=1}^{M-1}\sum_{n=0}^{N-1}10\log_{10}\rho_V(m,n)\right].\end{aligned}$$

These quantities were computed for $M = N = 16$ for $\lambda \in [0.1, 10]$ and the results are depicted in Figure 18. This figure shows that the maximum likelihood estimates have uniformly lower variance than the other estimates as guaranteed by the theory. Note that little is gained on the average using the maximum likelihood method for small values of $\lambda$, whereas a reduction of at least a few dB can be expected for $\lambda > 2$.

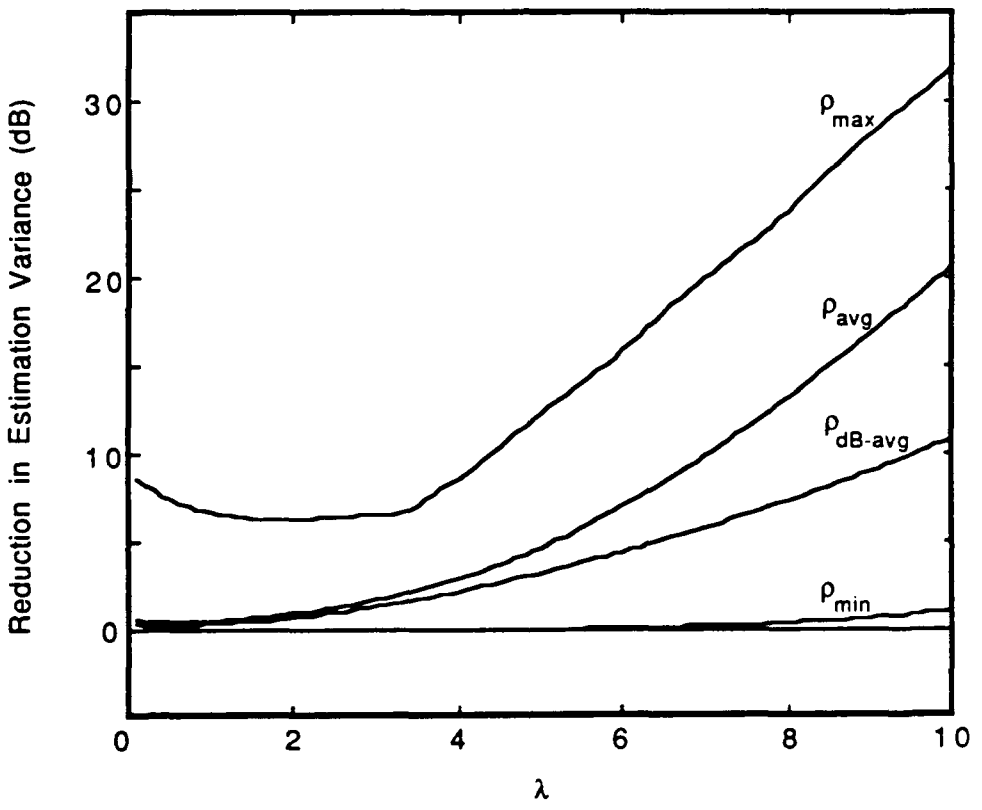



**Figure 18.** The maximum, average, and minimum reduction in the estimation variance obtained by using the maximum likelihood method instead of the biorthogonal function method to calculate the coefficients. The signal under consideration is a real signal with the Gabor representation with $M = N = 16$ using the exponential window function with parameter $\lambda$.

### 4.4.5 Detection for Finite Observation Intervals

In [4], the following detection problem was analyzed. Under the null hypothesis $H_0$ only white Gaussian noise is present. Under the alternate hypothesis $H_1$, the received signal consists of a transient signal together with additive white Gaussian noise. The transient is given by the representation in (41) where only the coefficients $\{A_{mn}; 0 \leq m \leq M-1, n \in \mathcal{N}\}$ can be non-zero, where $\mathcal{N}$ is a subset of $K$ integers in the range $[0, N-1]$. Basically, this detection problem amounts to deciding if a transient begins at a specified time with unknown frequency index. Since the values of the non-zero coefficients of the transient signal are assumed to be unknown, a generalized likelihood ratio test is derived which uses the coefficients derived from the BFM in [3,4]. The detector is summarized here. Reorder the vector $\mathbf{C}$ in (36) into the vector $\mathbf{x}$ so that

$$\mathbf{x} = \left[\mathbf{x}_1^T \ \mathbf{x}_2^T\right]^T$$

where $\mathbf{x_1}$ is the $(2M-1)K$ component vector corresponding to the nonzero coefficients in the transient signal and $\mathbf{x_2}$ is the $(2M-1)(N-K)$ component vector corresponding to the remaining coefficients. Let $\mathbf{\Lambda}$ be given by $\mathbf{\Lambda} = E[\mathbf{C}\mathbf{C}^T]$ which is defined in (37) and has been reordered to correspond to the new ordering of $\mathbf{C}$, and let $\mathbf{\Gamma} = \mathbf{\Lambda}^{-1}$. This reordering admits the following partitioning of $\mathbf{\Lambda}$

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{11} & \mathbf{\Lambda}_{12} \\ \mathbf{\Lambda}_{21} & \mathbf{\Lambda}_{22} \end{bmatrix}.$$

The test statistic derived from the generalized likelihood ratio test $T_1$ is given by [see 4 for details]

$$T_1 = \mathbf{x}^T \mathbf{\Lambda}^{-1} \mathbf{x} - \mathbf{x}_2^T \mathbf{\Lambda}_{22}^{-1} \mathbf{x}_2. \tag{62}$$

It is shown that under $H_0$, $T_1$ is chi-square distributed with $(2M-1)K$ degrees of freedom. Under $H_1$, $T_1$ has the noncentral chi-square distribution with $(2M-1)K$ degrees of freedom

and noncentrality parameter $\delta_1^2$ given by

$$\begin{aligned}\delta_1^2 &= \mathbf{u}_1^T \left(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right)^{-1} \mathbf{u}_1 \\ &= \mathbf{u}_1^T \Gamma_{11} \mathbf{u}_1\end{aligned} \tag{63}$$

where $\mathbf{u}$ is the vector of the true Gabor coefficients of the noise–free transient.

The same detection problem is now considered using the maximum likelihood approach in order to determine if the lower estimation variance of the maximum likelihood estimates translates into better detection performance as well. Again, a generalized likelihood ratio is used to determine the test statistic $T_2$

$$\begin{aligned}T_2 = &\max_{\{U_{mn},V_{mn};\ n\in\mathcal{N}\}} 2\log P\Big[\{r_t,\ t\in[0,N]\} \mid \{U_{mn},V_{mn}\}, U_{mn}=V_{mn}=0 \text{ for } n\notin\mathcal{N}\Big] \\ &\qquad - 2\log P\Big[\{r_t,\ t\in[0,N]\} \mid U_{mn}=V_{mn}=0\ \forall m,n\Big] \\ = &\frac{1}{N_0} \max_{\substack{\{U_{mn},V_{mn};\ n\in\mathcal{N}\} \\ U_{mn}=V_{mn}=0\ n\notin\mathcal{N}}} \Omega\left(\{U_{mn},V_{mn}\}\right).\end{aligned} \tag{64}$$

Again, reorder the coefficient vector W of (51) into a vector $\mathbf{y}$ so that

$$\mathbf{y} = \left[\mathbf{y}_1^T\ \mathbf{y}_2^T\right]^T$$

where $\mathbf{y}_1$ is the $(2m-1)K$ component vector corresponding to the non-zero coefficients of the noise free transient and $\mathbf{y}_2$ is the $(2m-1)(N-K)$ component vector corresponding to the other coefficients. Similarly, reorder the vector $\mathbf{T}$ of sufficient statistics in (53) into a vector $\mathbf{z} = [\mathbf{z}_1^T\ \mathbf{z}_2^T]^T$. Also, define a matrix $\mathbf{S}$ by reordering the matrix $(1/N_0)\mathbf{R}$ to match $\mathbf{y}$ and $\mathbf{z}$, where $\mathbf{R}$ is the signal-cross correlation matrix in (54). The matrix $\mathbf{S}$ admits the partition corresponding to $\mathbf{y}$ and $\mathbf{z}$,

$$\mathbf{S} = \begin{bmatrix}\mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22}\end{bmatrix}.$$

It follows from (55) and the fact that $\mathbf{y}_2 = \mathbf{0}$ that

$$\Omega\left(\{U_{mn},V_{mn}\}\right) = 2\mathbf{y}_1^T\mathbf{z}_1 - N_0\mathbf{y}_1^T\mathbf{S}_{11}\mathbf{y}_1 \tag{65}$$

and that $\mathbf{z}_1$ is the vector of the jointly sufficient statistics for this detection problem. The vector $\hat{\mathbf{y}}_1$ which maximizes (65) for $\mathbf{S}_{11}$ invertible is

$$\hat{\mathbf{y}}_1 = \frac{1}{N_0}\mathbf{S}_{11}^{-1}\mathbf{z}_1 \tag{66}$$

for which the achieved maximum is

$$\max_{\substack{\{U_{mn},V_{mn},\ n\in\mathcal{N}\}\\ U_{mn}=V_{mn}=0\ n\notin\mathcal{N}}} \Omega(\{U_{mn},V_{mn}\}) = \frac{1}{N_0}\mathbf{z}_1^T\mathbf{S}_{11}^{-1}\mathbf{z}_1 = N_0\hat{\mathbf{y}}_1^T\mathbf{S}_{11}\hat{\mathbf{y}}_1.$$

Thus, the test statistics is

$$T_2 = \hat{\mathbf{y}}_1^T\mathbf{S}_{11}\hat{\mathbf{y}}_1. \tag{67}$$

Note that the vector $\hat{\mathbf{y}}_1$ is not the same as the first partition of the overall maximum likelihood vector $\widehat{\mathbf{W}}_{\mathrm{MLE}}$ computed without restricting any coefficients to be zero. It is, however, the projection onto the subspace determined by the non-zero coefficients.

Since $\mathbf{S}_{11}$ is symmetric and positive semi-definite, by the Cholesky decomposition there exists a lower triangular matrix $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^T = \mathbf{S}_{11}$. In light of (57), $\mathbf{L}^T\hat{\mathbf{y}}_1$ is a vector of independent and identically distributed Gaussian random variables with zero mean and a variance of one. It follows that $T_2$ has the chi-square distribution with $(2M-1)K$ degrees of freedom under $H_0$. Under $H_1$, $T_2$ is noncentral chi-square with $(2M-1)K$ degrees of freedom and noncentrality parameter $\delta_2^2$ given by

$$\delta_2^2 = \mathbf{u}_1^T\mathbf{S}_{11}\mathbf{u}_1. \tag{68}$$

Using the approach in [4] another detector can be developed which uses the maximum likelihood estimates. To apply the same technique used to derive (62), let $\mathbf{w} = [\mathbf{w}_1^T\ \mathbf{w}_2^T]^T$ be the reordered vector of global maximum likelihood estimates $\widehat{\mathbf{W}}_{\mathrm{MLE}}$ in (58). From (57), the covariance matrix $\mathbf{\Sigma}$ of $\mathbf{w}$ is $\mathbf{\Sigma} = \mathbf{S}^{-1}$. The new detection statistic $T_3$ is then exactly

as in (62), with $\mathbf{w}$ and $\boldsymbol{\Sigma}$ replacing $\mathbf{x}$ and $\boldsymbol{\Lambda}$,

$$T_3 = \mathbf{w}^T \boldsymbol{\Sigma}^{-1} \mathbf{w} - \mathbf{w}_2^T \boldsymbol{\Sigma}_{22}^{-1} \mathbf{w}_2. \tag{69}$$

Using the matrix identity

$$\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \tag{70}$$

$T_3$ can be expressed in terms of the signal cross-correlation matrix by

$$T_3 = \mathbf{w}^T \mathbf{S} \mathbf{w} - \mathbf{w}_2^T \left( \mathbf{S}_{22} - \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \right) \mathbf{w}_2. \tag{71}$$

There are two things to note about $T_3$ from (69). First, this detection statistic is a function of the global maximum likelihood estimates, not the local ones as in $T_2$. Second, since $\mathbf{w} = (1/N_0)\mathbf{S}^{-1}\mathbf{z}$, $T_3$ is apparently a function of all the components of $\mathbf{z}$, not just the sufficient statistics which were shown to be $\mathbf{z}_1$ by (65).

Once again, $T_3$ has the chi-square distribution with $(2M-1)K$ degrees of freedom under $H_0$. Under $H_1$, $T_3$ is noncentral chi-square with $(2M-1)K$ degrees of freedom and noncentrality parameter $\delta_3^2$ given by

$$\begin{aligned} \delta_3^2 &= \mathbf{u}^T \left( \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right)^{-1} \mathbf{u} \\ &= \mathbf{u}_1^T \mathbf{S}_{11} \mathbf{u}_1. \end{aligned} \tag{72}$$

By comparing $\delta_2^2$ (68) and $\delta_3^2$ (72), it is apparent that $T_2$ and $T_3$ have identical distributions under $H_0$ and $H_1$, and, therefore, they have the same performance. It will now be shown that the two statistics are in fact equal.

**Proposition.** $T_2 = T_3$.

**Proof.** From (70) and using the identities $\mathbf{w} = (1/N_0)\mathbf{S}^{-1}\mathbf{z}$ and $\boldsymbol{\Sigma} = \mathbf{S}^{-1}$,

$$N_0^2 T_3 = \mathbf{z}^T \boldsymbol{\Sigma} \mathbf{z} - \left[ \boldsymbol{\Sigma}_{21} \mathbf{z}_1 + \boldsymbol{\Sigma}_{22} \mathbf{z}_2 \right]^T \boldsymbol{\Sigma}_{22}^{-1} \left[ \boldsymbol{\Sigma}_{21} \mathbf{z}_1 + \boldsymbol{\Sigma}_{22} \mathbf{z}_2 \right]$$

$$
\begin{aligned}
&= \mathbf{z}_1^T \boldsymbol{\Sigma}_{11} \mathbf{z}_1 + \mathbf{z}_1^T \boldsymbol{\Sigma}_{12} \mathbf{z}_2 + \mathbf{z}_2^T \boldsymbol{\Sigma}_{21} \mathbf{z}_1 + \mathbf{z}_2^T \boldsymbol{\Sigma}_{22} \mathbf{z}_2 \\
&\quad - \mathbf{z}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \mathbf{z}_1 - \mathbf{z}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{22} \mathbf{z}_2 - \mathbf{z}_2 \boldsymbol{\Sigma}_{22} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \mathbf{z}_1 - \mathbf{z}_2 \boldsymbol{\Sigma}_{22} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{22} \mathbf{z}_2 \\
&= \mathbf{z}_1^T \boldsymbol{\Sigma}_{11} \mathbf{z}_1 - \mathbf{z}_1^T \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \mathbf{z}_1 \\
&= \mathbf{z}_1^T \left( \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \right) \mathbf{z}_1 \\
&= \mathbf{z}_1^T \mathbf{S}_{11}^{-1} \mathbf{z}_1 \\
&= N_0^2 \mathbf{w}_1^T \mathbf{S}_{11} \mathbf{w}_1 \\
&= N_0^2 T_2. \qquad \blacksquare
\end{aligned}
$$

The implication of this result is that the test statistic $T_3$ in fact only uses the sufficient statistics and no more, even though it may appear otherwise.

The value of the second expression for the detection statistic can be seen from the following scenario. Suppose one is interested in testing several alternatives against the null hypothesis. Different alternatives may have one or more non-zero coefficients in common. Using the test statistic $T_2$, the local maximum likelihood estimates (LMLE) must be computed for each alternative, and the coefficients in common to more than one hypothesis will have different LMLE's. However, if the statistic $T_3$ is used instead, the global maximum likelihood estimates are computed for all the Gabor coefficients at once. These estimates are then processed according to (69) for each alternate hypothesis. This may be more efficient and also less ambiguous about estimates of coefficients in common to more than one alternative. Furthermore, it may be important from a data visualization standpoint to have estimates for all the Gabor coefficients, and thus the statistic $T_3$ is the natural choice.

Note that the noncentrality parameters $\delta_i^2$ in (63), (68), and (72) are generalized signal-to-noise ratios. They are linearly proportional to $\|\mathbf{u}\|^2$ and, since $\boldsymbol{\Gamma}_{11}$ in (63) and $\mathbf{S}_{11}$ in

(68) and (72) are inversely proportional to $N_0$, they are inversely proportional to $N_0$. Since the distributions of $T_1$ and $T_2$ (or $T_3$) have the same distribution under $H_0$ and, aside from the different noncentrality parameters, they also have the same distribution under $H_1$, it follows that the gain of one detection scheme over the other is fully determined by the noncentrality parameters. Accordingly, define the detection gain $G(\mathbf{u})$ in signal-to-noise ratio by

$$\begin{aligned} G(\mathbf{u}) &= \frac{\delta_1^2}{\delta_2^2} \\ &= \frac{\mathbf{u}^T \left(\Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}\right)^{-1} \mathbf{u}}{\mathbf{u}^T \left(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)^{-1} \mathbf{u}} \\ &= \frac{\mathbf{u}_1^T \Gamma_{11} \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{S}_{11} \mathbf{u}_1}. \end{aligned}$$

This gain represents the additional signal-to-noise ratio required for the detector $T_1$ of [4] to have the same probability of detection as the maximum likelihood detector $T_2$ or $T_3$ for a fixed probability of false alarm. It is easily shown to be independent of $\|\mathbf{u}\|^2$ and $N_0$.

Let $\mathbf{e}_i$ be the $(2M-1)N$ component vector such that $(\mathbf{e}_i)_i = 1$ and $(\mathbf{e}_i)_j = 0$ for $j \neq i$. Define the following quantities:

$$\begin{aligned} G_{\max} &= \max_{1 \le i \le (2M-1)N} G(\mathbf{e}_i) \\ G_{\min} &= \min_{1 \le i \le (2M-1)N} G(\mathbf{e}_i) \\ G_{\text{avg}} &= \frac{1}{(2M-1)N} \sum_{i=1}^{(2M-1)N} G(\mathbf{e}_i) \\ G_{\text{dB-avg}} &= \frac{1}{(2M-1)N} \sum_{i=1}^{(2M-1)N} 10 \log_{10} G(\mathbf{e}_i). \end{aligned}$$

These correspond to the maximum, minimum, and average gains for detecting a transient signal which has a single nonzero coefficient. In Figure 19, these quantities are shown for $M = N = 16$ and $\lambda \in [0.1, 10]$. These curves point to the same conclusion as the curves showing the reduction in estimation variance in Figure 18. Specifically, the maximum

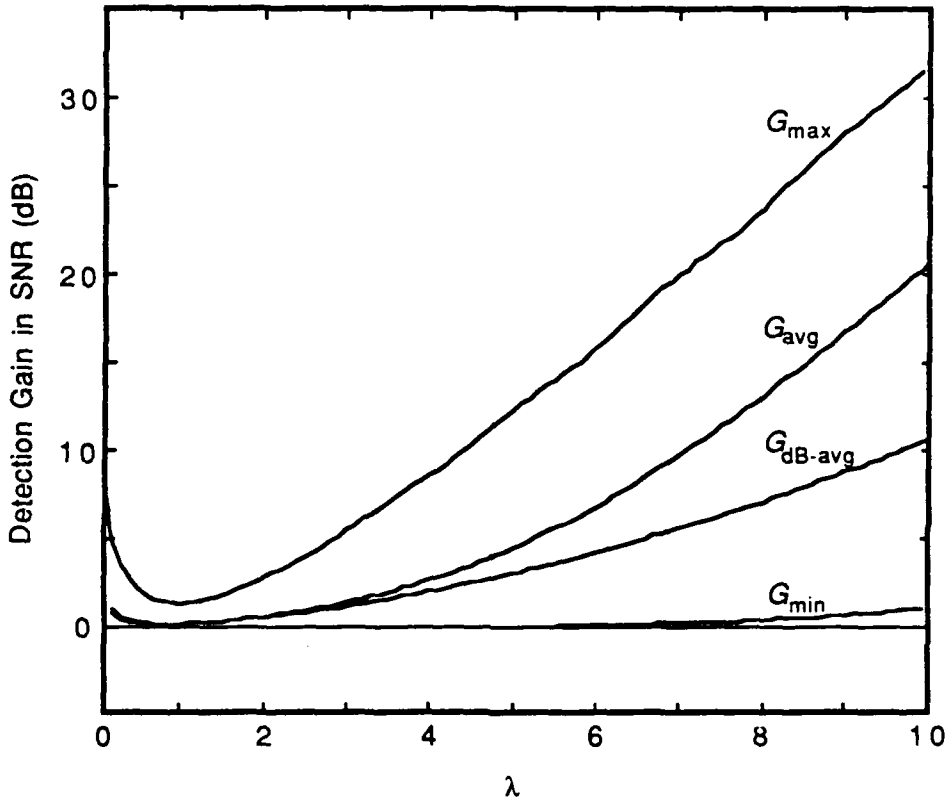



**Figure 19.** The maximum, average, and minimum gains in signal-to-noise ratio defined by the additional SNR required by the detector using the biorthogonal function method to attain the same performance as the detector using the maximum likelihood method. The signal under consideration is a real signal with a single non-zero coefficient in the Gabor representation with $M = N = 16$ using the exponential window function with parameter $\lambda$.

likelihood detector $T_2$ or $T_3$ is always better than the detector $T_1$, and on the average it will be several dB better for $\lambda > 2$.

## 4.5 The Exponential Window with Infinite Observation Intervals and Sequential Transient Detection

### 4.5.1 Introduction

The processing of a block of data representing a finite observation interval can be important once this block has been identified as warranting further analysis. However, transient signals have unknown arrival times, and in many applications data is arriving continuously. For these reasons, it is more practical to consider signal processing schemes which process the incoming data sequentially. Thus, when a transient actually arrives, it can be detected with little delay so that further processing and action can be taken.

With this philosophy in mind, the maximum likelihood technique for determining the Gabor coefficients using an exponential window is analyzed when the observation interval is the entire time line, that is $t \in \mathbb{R}$. At first glance, this would seem to make the processing unrealizable since the sufficient statistics now require matched filtering an infinite interval of data. However, it will be shown that the processing actually simplifies and is realizable when an infinite observation interval is considered. Moreover, the maximum likelihood processing results in the sequential processing of the data and in the sequential computation of the Gabor coefficients. Using the sequential estimation, a scheme is proposed and evaluated for the sequential detection of transient signals.

### 4.5.2 Maximum Likelihood Estimation of the Gabor Coefficients for Infinite Observation Intervals

There are two ways to analyze the computation of the maximum likelihood estimates of the Gabor coefficients when an infinite observation interval is considered. One uses filters and sequences and the other uses matrices and vectors. Because the finite observation interval problem has already been presented using matrices and vectors, it will be more intuitive to continue using this approach here. It should be noted, however, that both methods are equivalent. The matrices and vectors in this situation are "doubly-infinite" so that a matrix $\mathbf{A}$ has entries $A_{mn}$ for $m, n \in \mathbb{Z}$ and a vector $\mathbf{v}$ has entries $v_n$ for $n \in \mathbb{Z}$. Pictorially, the matrix can thought of as

$$\mathbf{A} = \begin{bmatrix} & \vdots & \vdots & \vdots & \\ \cdots & A_{-1,-1} & A_{-1,0} & A_{-1,1} & \cdots \\ \cdots & A_{0,-1} & A_{0,0} & A_{0,1} & \cdots \\ \cdots & A_{1,-1} & A_{1,0} & A_{1,1} & \cdots \\ & \vdots & \vdots & \vdots & \end{bmatrix}$$

and the vector can be symbolized by

$$\mathbf{v} = [\cdots v_{-1}\ v_0\ v_1 \cdots]^T .$$

The analysis in Section 4.4.3 concerning the sufficient statistics, the signal cross-correlations, and the functional form of the maximum likelihood estimates applies to the present situation as well, setting $I = \mathbb{R}$. Now, the sufficient statistics are given by

$$\begin{aligned} I_{mn} &= \int_{-\infty}^{\infty} g_e(t-n)\cos(2\pi mt)dr_t = \int_{n}^{\infty} g_e(t-n)\cos(2\pi mt)dr_t \\ Q_{mn} &= \int_{-\infty}^{\infty} g_e(t-n)\sin(2\pi mt)dr_t = \int_{n}^{\infty} g_e(t-n)\sin(2\pi mt)dr_t. \end{aligned} \tag{73}$$

These sufficient statistics are not realizable because they require observations from the semi-

infinite interval $[n,\infty)$. The signal cross-correlations in (48) are easily evaluated, yielding

$$r^{II}[(k,l),(m,n)] = 2\lambda^2 \exp\{-\lambda|n-l|\}\left[\frac{1}{4\lambda^2+4\pi^2(m-k)^2}+\frac{1}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{IQ}[(k,l),(m,n)] = \lambda \exp\{-\lambda|n-l|\}\left[\frac{2\pi(m-k)}{4\lambda^2+4\pi^2(m-k)^2}+\frac{2\pi(k+m)}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{QI}[(k,l),(m,n)] = \lambda \exp\{-\lambda|n-l|\}\left[\frac{2\pi(k-m)}{4\lambda^2+4\pi^2(m-k)^2}+\frac{2\pi(k+m)}{4\lambda^2+4\pi^2(m+k)^2}\right]$$

$$r^{QQ}[(k,l),(m,n)] = 2\lambda^2 \exp\{-\lambda|n-l|\}\left[\frac{1}{4\lambda^2+4\pi^2(m-k)^2}-\frac{1}{4\lambda^2+4\pi^2(m+k)^2}\right].$$

Once again, the signal cross-correlations factor into a product of two terms, one depending only on the difference in delay $|n-l|$ and the other depending only on the frequency indices $m$ and $n$. Thus, the signal cross-correlation matrix $\mathbf{R}$ factors into the Kronecker product $\mathbf{R} = \mathbf{R}_1 \otimes \mathbf{R}_2$ where $\mathbf{R}_1$ is the doubly-infinite Toeplitz matrix with entries

$$(\mathbf{R}_1)_{l,n} = \exp\{-\lambda|n-l|\}$$

and $\mathbf{R}_2$ is the same matrix as before (59). In order to compute the maximum likelihood estimates, the inverse of $\mathbf{R}$ and, hence, the inverse of $\mathbf{R}_1$ are required. The inverse of a doubly-infinite Toeplitz matrix can be found by computing the Z–transform of the sequence of coefficients in any column of the matrix, and finding the power series expansion of its reciprocal. Specifically,

$$\begin{aligned}
R_1(z) &= \sum_{n=-\infty}^{\infty} (\mathbf{R}_1)_{n,0} z^n \\
&= \sum_{n=-\infty}^{\infty} \exp\{-\lambda|n|\} z^n \\
&= \sum_{0}^{\infty} \exp\{-\lambda n\} z^{-n} + \sum_{0}^{\infty} \exp\{-\lambda n\} z^n - 1 \\
&= \frac{1}{1-\exp\{-\lambda\}z^{-1}} + \frac{1}{1-\exp\{-\lambda\}z} - 1 \\
&= \frac{1-\exp\{-2\lambda\}}{-\exp\{-\lambda\}z^{-1} + (1+\exp\{-2\lambda\}) - \exp\{-\lambda\}z}.
\end{aligned}$$

Thus, the reciprocal of the Z–transform is

$$\frac{1}{R_1(z)} = \frac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}} z^{-1} + \frac{1+\exp\{-2\lambda\}}{1-\exp\{-2\lambda\}} - \frac{\exp\{-\lambda\}}{1-\exp\{-2\lambda\}} z,$$

from which it follows that $\mathbf{R}_1^{-1}$ is symmetric tridiagonal Toeplitz with entries

$$\left(\mathbf{R}_1^{-1}\right)_{l,n} = \begin{cases} \dfrac{1+\exp\{-2\lambda\}}{1-\exp\{-2\lambda\}}, & \text{if } n = l \\ \dfrac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}}, & \text{if } |n-l| = 1 \\ 0, & \text{if } |n-l| > 1. \end{cases} \tag{74}$$

The structure of the vector of sufficient statistics $\mathbf{T}$ which is composed of the doubly-infinite sequence of $(2M-1)$ component vectors $\mathbf{T}_n$ in (52) is

$$\mathbf{T} = \left[\cdots \mathbf{T}_{-1}^T \; \mathbf{T}_0^T \; \mathbf{T}_1^T \cdots\right]^T.$$

Define a doubly-infinite sequence of $(2M-1)$ component vectors $\mathbf{t}_n$ by

$$\mathbf{t}_n = \frac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}}\mathbf{T}_{n-1} + \frac{1+\exp\{-2\lambda\}}{1-\exp\{-\lambda\}}\mathbf{T}_n + \frac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}}\mathbf{T}_{n+1}.$$

The sequence $\{\mathbf{t}_n\}$ is the result of filtering the sufficient statistics with $\mathbf{R}_1^{-1}$. In keeping with the structure of $\mathbf{T}_n$, let the components of $\mathbf{t}_n$ be denoted $i_{mn}$ and $q_{mn}$ according to

$$\mathbf{t}_n = \left[i_{0,n} \; i_{1,n} \cdots i_{M-1,n} \; q_{1,n} \cdots q_{M-1,n}\right]^T.$$

Define the function $h(t)$ by

$$\begin{aligned} h(t) &= \frac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}}g_e(t+1) + \frac{1+\exp\{-2\lambda\}}{1-\exp\{-\lambda\}}g_e(t) + \frac{-\exp\{-\lambda\}}{1-\exp\{-2\lambda\}}g_e(t-1) \\ &= \frac{\sqrt{2\lambda}}{1-\exp\{-\lambda\}}\exp\{-\lambda t\}\left[-\exp\{-2\lambda\}u(t+1) + (1+\exp\{-2\lambda\})\,u(t) - u(t-1)\right]. \end{aligned}$$

Note that $h(t)$ is zero outside the interval $[-1,1]$. It can easily be checked that $i_{mn}$ and $q_{mn}$ are given by

$$i_{mn} = \int_{-1}^{1} h(t-n)\cos(2\pi m t)dr_t$$

$$q_{mn} = \int_{-1}^{1} h(t-n)\sin(2\pi m t)dr_t.$$

These equations are reminiscent of the filtering in the biorthogonal function method in (34). The filter $h(t)$ is shown in Figure 20 for $\lambda = 0.5$ and $\lambda = 4$ for comparison to the biorthogonal function in Figure 10 and Figure 11. It is apparent that $h(t)$ performs matched filtering, in essence, by amplifying regions with large signal values and attenuating areas with small signal values. This is in contrast with the biorthogonal function which does the reverse. Also, the calculation of the statistics $\{i_{mn}, q_{mn}\}$ is realizable and does not require the calculation of $\{I_{mn}, Q_{mn}\}$. A further simplification of the computation can be made by setting

$$\tilde{i}_{mn} = \frac{\sqrt{2\lambda}}{1 - \exp\{-\lambda\}} \int_0^1 \exp\{-\lambda t\} \cos(2\pi m t) dr_{t+n}$$

$$\tilde{q}_{mn} = \frac{\sqrt{2\lambda}}{1 - \exp\{-\lambda\}} \int_0^1 \exp\{-\lambda t\} \sin(2\pi m t) dr_{t+n}$$

so that

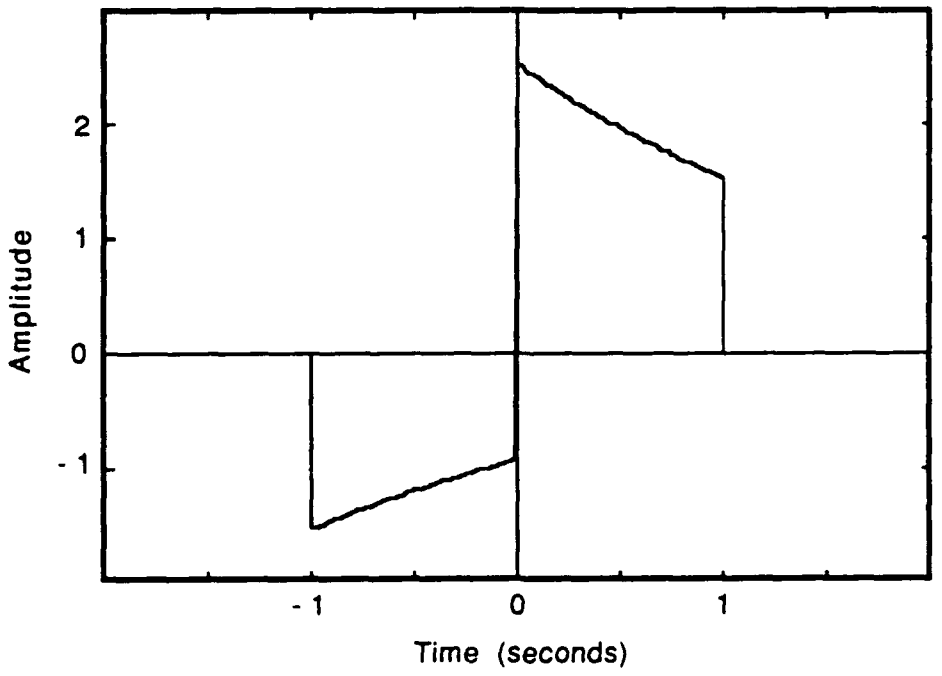
$$\begin{aligned} i_{mn} &= (1 + \exp\{-2\lambda\})\tilde{i}_{m,n} - \exp\{-2\lambda\}\tilde{i}_{m,n-1} \\ q_{mn} &= (1 + \exp\{-2\lambda\})\tilde{q}_{m,n} - \exp\{-2\lambda\}\tilde{q}_{m,n-1}. \end{aligned} \tag{75}$$

The $(2M-1)$ component of maximum likelihood estimates of the coefficients with time index $n$, $\widehat{\mathbf{W}}_n$ with form (50), is given by

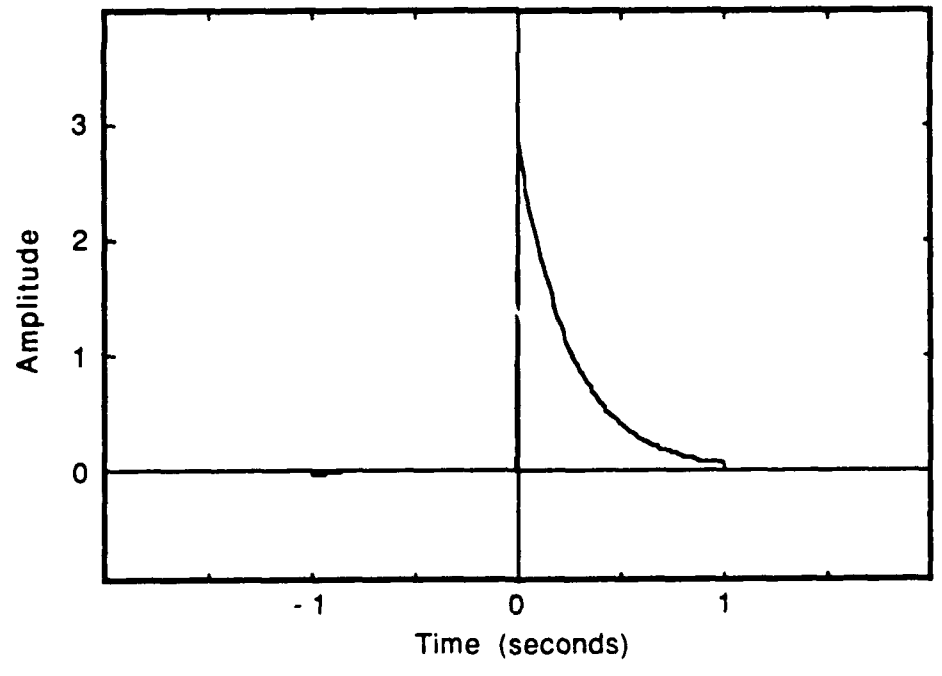
$$\widehat{\mathbf{W}}_n = \mathbf{R}_2^{-1}\mathbf{t}_n.$$

Thus, the calculation of the coefficient estimates for the time index $n$ depends only on the observations from the interval $[n-1, n+1]$. This yields the desired sequential implementation.

Because a simple analytic form is available for $\mathbf{R}_1^{-1}$, the estimator variance of the two schemes can be analytically compared when $M = 1$, that is when only baseband transients are considered. Then, $\mathbf{R}_2$ reduces to a scalar. The variances are found from (57),(59),(61),




(a)




(b)

**Figure 20.** The filter $h(t)$ used in the sequential computation of the maximum likelihood coefficient estimates for (a) $\lambda = 0.5$ and (b) $\lambda = 4$.

and (74) to be

$$\begin{aligned} \mathrm{Var}\left[\widehat{A}_n^{\mathrm{BFM}}\right] &= \frac{N_0}{2\lambda^2}\sinh(2\lambda), \\ \mathrm{Var}\left[\widehat{A}_n^{\mathrm{MLE}}\right] &= N_0 \coth(\lambda) \end{aligned} \tag{76}$$

where $A_n$ is the coefficient of the baseband pulse with delay $n$. The variances in (76) are plotted as a function of $\lambda$ in Figure 21. The figure shows that the maximum likelihood estimate has uniformly lower variance than the BFM estimate, and the reduction in estimation variance is most pronounced for large $\lambda$.

### 4.5.3 Sequential Transient Detection

Using the results of the previous section, a detection scheme is proposed which uses the maximum likelihood estimates of the coefficients computed by the sequential algorithm. Specifically, a quadratic form of the estimated Gabor coefficients at each time index and frequency index is compared to a threshold. If one of the test statistics exceeds the threshold, the presence of a transient is declared. This test also provides transient classification because the positive test yields the transient frequency and delay. The details are as follows.

Let $T_{mn}$ be the test statistic for a transient with frequency index $m$ and delay $n$. It is defined as follows. For $m = 0$, reorder the signal cross-correlation matrix $\mathbf{S}$ so that

$$(\mathbf{S})_{1,1} = \frac{1}{N_0} r^{II}[(0,0),(n,n)].$$

Then, $T_{0,n}$ is simply

$$T_{0,n} = \left(\mathbf{S}^{-1}\right)_{1,1} \widehat{U}_{0,n}^2.$$

For the other frequency indices, a transient with a particular frequency index and phase shift would show up in both $\widehat{U}_{mn}$ and $\widehat{V}_{mn}$ so that a quadratic form is desired that combines the two. Fix $m$ and $n$. Reorder the signal cross-correlation matrix so that the $\mathbf{S}_{11}$ is the

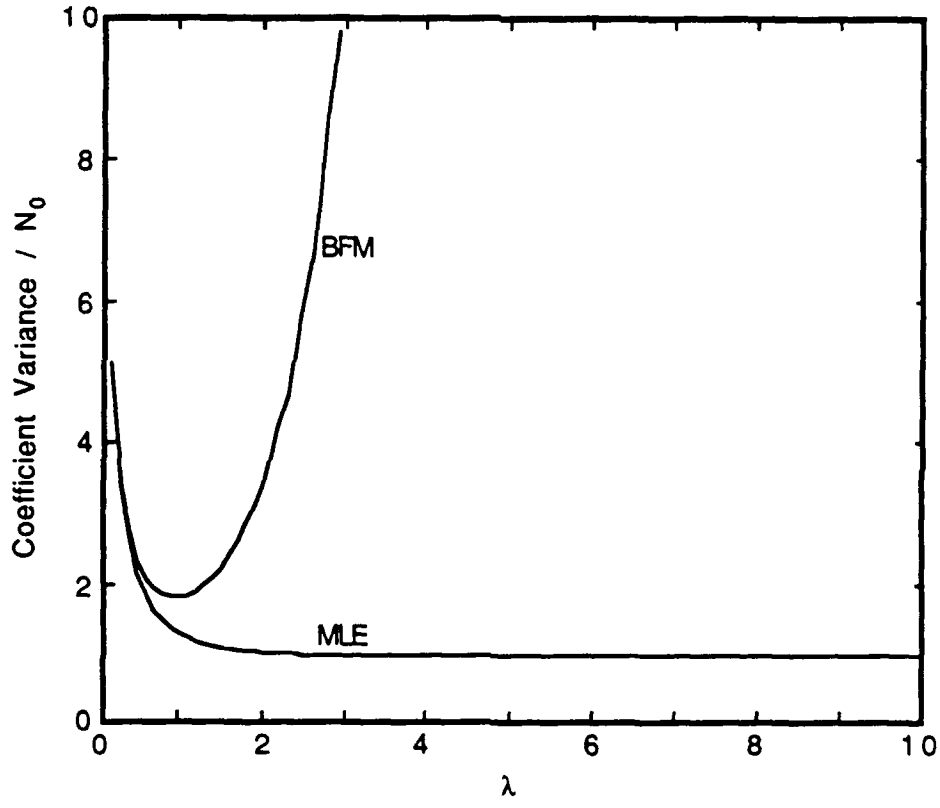



**Figure 21.** The variances of the estimates of the coefficients by the sequential processing implementing the biorthogonal function method and the maximum likelihood method. The signal under consideration is a real signal with the Gabor representation with $M = 1$ (baseband only transients) using the exponential window function with parameter $\lambda$.

$2 \times 2$ matrix corresponding to the pair $(m, n)$, that is

$$\mathbf{S}_{11} = \frac{1}{N_0} \begin{bmatrix} r^{II}[(m,m),(n,n)] & r^{IQ}[(m,m),(n,n)] \\ r^{QI}[(m,m),(n,n)] & r^{QQ}[(m,m),(n,n)] \end{bmatrix}.$$

The covariance matrix of $\widehat{\mathbf{W}}_{mn} = \left[\hat{U}_{mn}\ \hat{V}_{mn}\right]^T$ is then $\mathbf{\Sigma}_{11} = \left(\mathbf{S}^{-1}\right)_{11}$. It is easy to show that the generalized likelihood ratio to test the hypothesis "$\widehat{\mathbf{W}}_{mn} \sim N(\mathbf{u}, \mathbf{\Sigma}_{11})$," where $\mathbf{u} \in \mathbb{R}^2$ is an arbitrary non-zero vector, versus the hypothesis "$\widehat{\mathbf{W}}_{mn} \sim N(\mathbf{0}, \mathbf{\Sigma}_{11})$" is simply

$$T_{mn} = \widehat{\mathbf{W}}_{mn}^T \mathbf{\Sigma}_{11}^{-1} \widehat{\mathbf{W}}_{mn},$$

which is the test statistic for a transient with frequency index $m \neq 0$ and delay $n$.

Note that this test statistic was developed using the maximum likelihood estimates as the observations, not from the original observations. In terms of the original observations, this test corresponds to the generalized likelihood ratio for testing "$\{U_{kl}, V_{kl}$ arbitrary $k, l \in \mathbb{Z}\}$" against "$\{U_{kl}, V_{kl}$ arbitrary $k \neq m,\ l \neq n,\ U_{mn} = V_{mn} = 0\}$" as is shown below. The generalized likelihood $l(\mathbf{z})$ ratio reduces to

$$\begin{aligned} l(\mathbf{z}) &= \max_{\mathbf{y}} \left(2\mathbf{y}^T\mathbf{z} - N_0\mathbf{y}^T\mathbf{S}\mathbf{y}\right) - \max_{\mathbf{y}_2} \left(2\mathbf{y}_2^T\mathbf{z} - N_0\mathbf{y}_2^T\mathbf{S}\mathbf{y}_2\right) \\ &= \frac{1}{N_0}\mathbf{z}^T\mathbf{S}^{-1}\mathbf{z} - \frac{1}{N_0}\mathbf{z}_2^T\mathbf{S}_{22}^{-1}\mathbf{z}_2 \\ &= \frac{1}{N_0}\mathbf{z}^T\mathbf{\Sigma}\mathbf{z} - \frac{1}{N_0}\mathbf{z}_2^T\left(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)\mathbf{z}_2 \\ &= \frac{1}{N_0}\left(\mathbf{z}_1^T\mathbf{\Sigma}_{11}\mathbf{z}_1 + 2\mathbf{z}_1^T\mathbf{\Sigma}_{12}\mathbf{z}_2 + \mathbf{z}_2\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\mathbf{z}_2\right) \\ &= \frac{1}{N_0}\left(\mathbf{\Sigma}_{11}\mathbf{z}_1 + \mathbf{\Sigma}_{21}\mathbf{z}_2\right)^T\mathbf{\Sigma}_{11}^{-1}\left(\mathbf{\Sigma}_{11}\mathbf{z}_1 + \mathbf{\Sigma}_{21}\mathbf{z}_2\right) \\ &= \widehat{\mathbf{W}}_{mn}^T\mathbf{\Sigma}_{11}^{-1}\widehat{\mathbf{W}}_{mn} \\ &= T_{mn}. \end{aligned}$$

This test is in contrast with the fixed sample size test in Section 4.4.5 because it was assumed there that all the coefficients are zero except possibly the ones of interest. Here, where there

is an infinite number of coefficients, that is not a realistic assumption. Thus, the coefficients of interest are tested for being non-zero while not restricting the other coefficients to be zero.

The sequential detection algorithm with thresholds $h_0$ and $h$ is simply

$$\begin{aligned} T_{0,n} > h_0 &\Longrightarrow \text{Declare transient present in bin } (0,n) \\ T_{1,n} > h &\Longrightarrow \text{Declare transient present in bin } (1,n) \\ \vdots \qquad &\qquad \vdots \\ T_{M-1,n} > h &\Longrightarrow \text{Declare transient present in bin } (M-1,n) \\ &\text{Continue, } n = n+1. \end{aligned}$$

Equivalently, for each $n$, the maximum of the statistics $\{T_{0,n}/h_0;\ T_{mn}/h,\ m = 1,\ldots M-1\}$ could be compared to a threshold of one. This could be generalized to detect more complicated transients which involve several damped sinusoids by comparing some function of the $k$ out of $M$ largest order statistics to a threshold.

The distribution of the test statistic $T_{0,n}$ is chi-square with one degree of freedom when the corresponding coefficient $A_{0,n}$ is zero. It is non-central chi-square with one degree of freedom with noncentrality parameter $\delta^2_{0,n}$ given by

$$\delta^2_{0,n} = A^2_{0,n} \left(\mathsf{S}^{-1}\right)_{0,n}$$

when $A_{0,n} \neq 0$. The distribution of the test statistic $T_{mn}$ for $m \neq 0$ is chi-square with two degrees of freedom when no transient is present and is non-central chi-square with two degrees of freedom and noncentrality parameter $\delta^2_{mn}$ given by

$$\delta^2_{mn} = \mathbf{u}^T_{mn} \Sigma^{-1}_{11} \mathbf{u}_{mn},$$

where $\mathbf{u}_{mn} = [U_{mn}\ V_{mn}]^T$ is the true coefficient vector.

For a given threshold $h$, the threshold $h_0$ can be set according to

$$P\{\chi_1^2 > h_0\} = P\{\chi_2^2 > h\}, \tag{77}$$

where $\chi_1^2$ and $\chi_2^2$ are chi-square random variables with one and two degrees of freedom. This means that the probability of false alarm $P_{fa}$ for the decision for each coefficient, which is given by the expressions in (77), will be the same for all $m$. In order to determine the mean time between false alarms, the following convention is assumed. For each fixed $n$, if one or more of the test statistics exceed the threshold, it is considered as a single false alarm. If an alarm, false or otherwise, is detected at time $n$, then any alarm at time $n+1$ is not counted as a separate alarm. Using these conventions, and the fact that the maximum likelihood coefficients for bins $(k,l)$ and $(m,n)$ are independent for $|n-l| > 1$ as a consequence of (74), the mean time between false alarms $T_{fa}$ is bounded by

$$\frac{2}{MP_{fa}} \leq T_{fa} \leq \frac{2}{P_{fa}}.$$

The probability of detection $P_d$ is given by

$$P_d = \begin{cases} P\{\chi_1^2(\delta_{0,n}^2) > h_0\}, & \text{for } m = 0 \\ P\{\chi_2^2(\delta_{m,n}^2) > h\}, & \text{for } m \neq 0, \end{cases}$$

where $\chi_i^2(\delta^2)$ is a non-central chi-square random variable with $i$ degrees of freedom and noncentrality parameter $\delta^2$.

## 4.6 Conclusion

In this chapter, the Gabor representation was discussed because of its desirable localization properties in both time and frequency. Two methods of computing and estimating the Gabor coefficients were examined. The maximum likelihood estimates and detectors based on them were shown to be superior to those obtained by the more traditional approach using the biorthogonal function. Using an exponential window function, the Gabor representation was applied to the detection of transient signals. A simple sequential detection procedure was given to detect transient signals using the maximum likelihood estimates.

There are several avenues for further research. Throughout the chapter it has been assumed that the delays and frequencies could only take on a discrete set of values indexed by the integers. It would be necessary for a real application to determine the performance when this assumption is not valid. Also, it was assumed that $\tau^2$ for the Gaussian window and $\lambda$ for the exponential window were known a priori. This may not be the case in reality. Thus, the effect of the window mismatch could be examined. Perhaps an optimal window function for a particular application could be designed to minimize the mismatching effects, or a bank processors could be used to implement several different window functions simultaneaously. Finally, detection procedures using the Gabor representation could be compared to more traditional signal processing techniques and representations such as Prony's method and the short-time Fourier Transform.

## 4.7 References


[1] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. of the IEEE*, vol. 68, no. 4, pp. 538–539, April 1980.

[2] ——,"A sampling theorem for the complex spectrogram, and Gabor's expansion of a signal in Gaussian elementary signals," *Optical Engineering*, vol. 20, no. 4, pp. 594–598, July 1981.

[3] B. Friedlander, "Transient signal detection techniques," draft of paper for *U.S. Navy Journal on Underwater Acoustics*, October 13, 1989.

[4] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-37, no. 2, pp. 169–180, February 1989.

[5] D. Gabor, "Theory of Communication," *Proc. of the IEEE*, vol. 93, pt. III, pp. 429–457, 1946.

[6] C. W. Helstrom, "An expansion of a signal in Gaussian elementary signals," *IEEE Trans. on Info. Theory*, vol. IT-12, pp. 81–82, January 1966.

[7] A. J. E. M. Janssen, "Gabor representation and Wigner distribution of signals," *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, no. 41B.2, 1984.

[8] L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis*, Groningen, The Netherlands: P. Noordhoff, Ltd., Section 1.2, 1958.

[9] E. L. Lehmann, *Theory of Point Estimation*, New York: Wiley, 1983.

[10] D. G. Luenberger, *Optimization by Vector Space Methods*, New York: Wiley, 1969.

[11] S. L. Marple, Jr., "Transient signal enhancement by linear predictive eigenanalysis," *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, no. U2.8, pp. 2925–2928, 1988.

[12] L. K. Montgomery, Jr., "A generalization of the Gabor–Helstrom transform," *IEEE Trans. Info. Theory*, vol. IT-13, pp. 344–345, April 1967.

[13] C. H. Page, "Instantaneous power spectra," *J. Applied Physics*, vol. 23, no. 1, pp. 103–106, January 1952.

[14] A. J. Poggio, *et al.*, "Evaluation of a processing technique for transient data," *IEEE Trans. Antennas and Propagation*, vol. AP-26, no. 1, pp. 165–173, January 1978.

[15] M. Porat and Y. Y. Zeevi, "The generalized Gabor scheme of image representation in biological and machine vision," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 452–467, July 1988.

[16] M. Porat and Y. Y. Zeevi, "Localized texture processing in vision: analysis and synthesis in the Gaborian space," *IEEE Trans. Biomedical Engineering*, vol. 36, no. 1, pp. 115–129, January 1989.

[17] M. B. Priestly, "Evolutionary spectra and non-stationary processes," *J. Royal Statist. Soc. B*, vol. 27, no. 2, pp. 204–237, 1965.

[18] M. L. Van Blaricum and R. Mittra, "Problems and solutions associated with Prony's method for processing transient data," *IEEE Trans. Antennas and Propagation*, vol. AP-26, no. 1, pp. 174–182, January 1978.

[19] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Info. Theory*, vol. IT-32, no. 1, pp. 85–96, January 1986.

# — 5 —

# The Optimal Sample Size for Transient Detection

## 5.1 Introduction

Often when sequential data processing algorithms are implemented in hardware, the data are grouped into fixed size blocks to facilitate the implementation. An example of this type of processing is a hardware implementation of the Fast Fourier Transform which might be used to sequentially compute short time Fourier Transforms as data arrive. The same approach could be used to compute the Gabor coefficients discussed in Chapter 4. In that context, the transient model prescribed the optimal block size for the processing. The problem considered in this chapter is to find the optimal sample size or integration period for the detection of a transient which arrives at an unknown time.

The approach using fixed sample size tests in place of a fully sequential algorithm was first considered for the quickest detection problem [2,3,4]. In [4], Shiryaev considered the quickest detection problem for detecting an abrupt change in the drift of a Brownian motion. In the fixed sample size approach which he called the Neyman-Pearson method, the test statistics were based on observations from consecutive intervals of fixed length. He

then considered the interval length that would minimize the mean delay in the detection of the change for a fixed mean time between false alarms when the change is assumed to begin with uniform distribution during the interval in which it actually occurs. Pelkowitz and Schwartz [2] and Pelkowitz [3] considered the quickest detection problem in discrete time. By using the Central Limit Theorem for large sample sizes, they were able to find the optimal sample sizes for Gaussian and non-Gaussian samples. Two different criteria were examined, the mean criterion and the minimax criterion. The mean criterion is the same one that Shiryaev used, i.e., minimize the mean delay for a fixed mean time between false alarms, when the change time is equally likely to occur any time during the block. The minimax criterion attempts to minimize the expected delay for the least favorable change time for a fixed mean time between false alarms. The least favorable change time is defined to be the time during the integration period which yields the largest expected delay in detection. Using an asymptotic analysis, a procedure is presented to determine the optimal sample size and the detection threshold for these criteria.

In this chapter, the question of the optimal sample size for discrete-time observations or the optimal integration period for continuous-time observations will be examined in the context of transient signal detection. In order to obtain tractable solutions for the discrete-time case, it is assumed that either the observations are Gaussian or that the transient lasts for a large number of samples so that the Central Limit Theorem (CLT) can be applied. As in [2,3], the discrete-time detection scheme is based on fixed sample size test statistics $S_n$ of the form

$$S_n = \sum_{m=nT}^{n(T+1)-1} z_m.$$

Here, $T$ is the integer-valued sample size and $z_n$ is some function of the raw observations $\{x_n\}$. Candidate functions include the log-likelihood ratio or the squared magnitude of a DFT coefficient. These statistics are computed sequentially and a detection is declared

when a threshold $h$ is exceeded. In continuous time, some form of the CLT could be used for non-Gaussian observations, but this raises many technical issues such as the definition of the continuous-time integral. In order not to complicate the essential problem being investigated, the continuous-time observations are assumed to come from a white Gaussian process. The detection scheme in continuous time is based on test statistics $S_n$ of the form

$$S_n = \int_{(n-1)T}^{nT} x(t)dt.$$

Here, $T$ is the real-valued integration period and $x(t)$ represents the observations. Again, these statistics are compared to a threshold $h$ to declare a detection.

The optimal sample size and integration period for the detection of transient signals will be examined with the optimization carried out with respect to both the mean criterion and the minimax criterion. The performance criteria are the same as those in Chapter 3, the mean time between false alarms and the probability of detection. Thus, instead of minimizing the expected delay as is done for the quickest detection problem, the probability of detection is maximized according to one of the two optimality criteria for a fixed mean time between false alarms.

In addition to the false alarm constraint in the optimization problem, an additional constraint is instituted to achieve desirable solutions. That constraint specifies that the admissible set of sample sizes for the optimization procedure is an interval restricted to some neighborhood of the transient duration. The are two reasons to do this. First, this ensures there will be relatively little delay in detecting the transient. Second, if this constraint were not enforced, the optimization would yield a degenerate and useless solution as can be seen as follows. The best performance one could hope for is a very large mean time between false alarms and a probability of detection near one, but this is actually very easy obtain.

Consider a detection scheme with the integration period equal to the desired mean time between false alarms and a detector which always declares a transient present. This detector has exactly the desired mean time between false alarms with a probability of detection of one and yet it provides no useful information.

The outline for the remainder of this chapter is as follows. In Section 2, the exact probabilistic model considered is presented and the necessary equations for the two optimality criteria are derived. Numerical results are discussed in Section 3. Section 4 considers the application of the detection scheme to narrowband signals and conclusions are presented in Section 5.

## 5.2 The Transient Detector and the Optimality Criterion

We first consider the discrete-time situation. The statistical model for the transient is specified in a manner similar to the model in [2,3]. Assume there exists a random sequence $y_n$ such that:

i) $y_n$ is a strict-sense stationary process with $0 < E[|y_n|^m] < \infty$ for $m \leq 3$.

ii) $y_n$ is $L$–dependent, i.e., there exists an integer $L$ such that if the event $A$ depends on $\{y_m, m \leq n\}$ and an event $B$ depends on $\{y_m, m > L+n\}$, then $A$ and $B$ are independent.

Assume the transient lasts $\tau$ samples and it begins at the $m$'th sample of the $n$'th block, i.e., at time $t_0 = T(n-1)+m$. Then, the observations in terms of the $\{z_n\}$ are given by

$$z_n = \begin{cases} y_n, & \text{for } n < t_0 \\ y_n + A, & \text{for } t_0 \leq n \leq t_0 + \tau - 1 \\ y_n, & \text{for } n \geq t_0 + \tau. \end{cases}$$

Here, $A$ is a positive constant. This model is not as complicated as the one used in the disorder problem in [2,3]. The model used there allowed for a change in the second order statistics in addition to a change in mean. In that context, the covariance structures before and after the change were made to approach each other asymptotically. A local asymptotic approach such as this is valid for the disorder problem because, once the change time occurs, there is an arbitrarily long time to detect the change since the change lasts forever. Thus, even extremely weak signals could be detected by integrating and/or waiting long enough. In the transient signal detection problem where the signal has finite duration, this small signal approach is not realistic, and the problem becomes intractable using the more general model.

As pointed out in [2,3], assumption ii) implies that $S_n$ and $S_{n+m}$ are independent for $T \geq L$ and $m \geq 2$. Moreover, if $T \gg L$, the $L$ samples at the beginning and end of the sum will contribute a negligible amount to the overall sum so that all the $S_n$ will essentially be mutually independent. It is proved in [3] that as $T \rightarrow \infty$ the $\{S_n\}$ tend to mutually independent Gaussian random variables using the CLT and assumptions i) and ii). We will use this fact for large $T$ with $T \gg L$.

As in [2,3], we define the following parameters:

$$\mu_0 = E[y_n]$$

$$\sigma_0^2 = \mathrm{Var}[y_n]$$

$$r_0(m) = \frac{1}{\sigma_0^2} E[(y_n - \mu_0)(y_{n+m} - \mu_0)] = r_0(-m)$$

$$\gamma_0 = \sum_{m=-L}^{L} r_0(m) = 1 + 2\sum_{m=1}^{L} r_0(m)$$

$$\sigma^2 = \gamma_0 \sigma_0^2$$

$$\beta = \frac{A}{\sqrt{\gamma_0}\sigma_0} = \frac{A}{\sigma}.$$

The effective noise power $\sigma^2$ and the effective single sample signal-to-noise ratio (SNR) $\beta$ take into account the effect of the noise correlation.

For the continuous-time model, the observed random process $x(t)$ is defined by

$$x(t) = A\,[u(t - t_0) - u(t - t_0 - \tau)] + n(t), \tag{1}$$

where $A$ is a positive constant, $u(t)$ is the unit step function, $t_0$ is the unknown arrival time, $\tau$ is the transient duration which is assumed known, and $n(t)$ is a white Gaussian noise process with mean $\mu_0$ and autocorrelation function $E[(n(t) - \mu_0)(n(t + s) - \mu_0)] = \sigma^2\delta(s)$. The SNR is defined by $\beta = A/\sigma$. Setting $\tau = \infty$ in (1) results in the quickest detection problem that Shiryaev investigated [4].

Note that the same variables have been used for the discrete-time and continuous-time cases, namely the test statistics $S_n$, the sample size or integration period $T$, the arrival time of the transient $t_0$, the transient duration $\tau$, the mean of the noise $\mu_0$, the effective noise variance $\sigma^2$, and the effective SNR $\beta$. In the remainder of this chapter both interpretations of the variables apply unless otherwise stated.

When the transient overlaps the $n$'th block of samples or the $n$'th integration period, the distribution of the test statistic $\{S_n\}$ is

$$S_n \sim \mathcal{N}(A\tau_n + \mu_0 T, T\sigma^2),$$

where $\tau_n$ represents the duration (in samples or seconds) of the period of overlap of the transient in the $n$'th block. To derive this for the discrete time case, it is assumed that $T$ is large and $T \gg L$ [2,3]. The random variables $\{S_n\}$ are mutually independent.

When $\tau_n = 0$, such as before the transient arrives, the test statistic is distributed

according to $S_n \sim N(\mu_0 T, T\sigma_2)$. Thus, the probability $\alpha$ that a false alarm results from $S_n$ is given by

$$\alpha = P\{S_n > h | \tau_n = 0\} = 1 - \Phi\left(\frac{h - \mu_0 T}{\sqrt{T\sigma^2}}\right),$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. Thus, on the average there will be one false alarm every $1/\alpha$ integration periods so that the mean time between false alarms $M_F$ is

$$M_F = \frac{T}{1 - \Phi\left(\frac{h - \mu_0 T}{\sqrt{T\sigma^2}}\right)}.$$

It will be convenient to define a normalized threshold $\lambda$ given by

$$\lambda = \Phi^{-1}\left(\frac{T}{M_F}\right) - \frac{\mu_0 T}{\sqrt{T\sigma^2}} = \frac{-h}{\sqrt{T\sigma^2}}. \tag{2}$$

The probability of detection is a function of the true arrival time $t_0$, but it suffices to consider the arrival time modulo $T$ so that it is assumed that $t_0$ lies in the range $0 < t_0 \leq T$. There are two cases to consider: $T \geq \tau$ and $T < \tau$. For the first case, the transient can either overlap with one or two blocks as shown in Figure 1. The transient is completely contained in one integration period if $t_0 \leq T - \tau$. If $t_0 > T - \tau$, then the transient overlaps with two blocks. The probability of detection $P_d(t_0)$ is easily found to be

$$P_d(t_0) = \begin{cases} \Phi\left(\frac{\beta\tau}{\sqrt{T}} + \lambda\right), & \text{if } t_0 < T - \tau \\ \Phi\left(\frac{\beta(T-t_0)}{\sqrt{T}} + \lambda\right) + \left[1 - \Phi\left(\frac{\beta(T-t_0)}{\sqrt{T}} + \lambda\right)\right] \Phi\left(\frac{\beta(\tau-T+t_0)}{\sqrt{T}} + \lambda\right), & \text{if } t_0 \geq T - \tau. \end{cases} \tag{3}$$

When $T \leq \tau$, the transient overlaps with at least two intregration periods as shown in Figure 2. As seen in the figure, there are blocks on either end of the transient that contain a mixture of noise only and transient plus noise parts, and there are a number of blocks in the center which are completely characterized by the transient plus noise situation. The

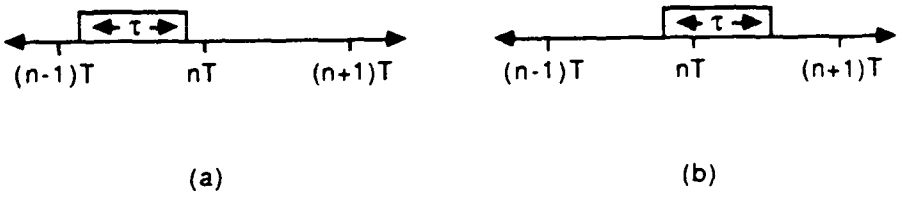



**Figure 1.** When $T > \tau$, the transient can overlap (a) one or (b) two integration periods.

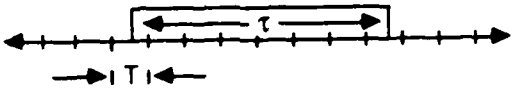



**Figure 2.** An example of how the transient overlaps the integration periods for $T < \tau$.

number $k(t_0)$ of the center-type blocks for the continuous-time case can be found to be

$$k(t_0) = \begin{cases} \lfloor \tau/T \rfloor - 1, & \text{if } 0 < t_0 < T(\lfloor \tau/T \rfloor + 1) - \tau \\ \lfloor \tau/T \rfloor, & \text{if } T(\lfloor \tau/T \rfloor + 1) - \tau \leq t_0 \leq T, \end{cases} \tag{4}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to $x$. For the discrete-time case, the two regions of support for $k(t_0)$ may differ from those in (4) by one sample, but this will be insignificant when $T$ is large. The probability of a detection resulting from the center-type blocks $P_d^c(t_0)$ is one minus the probability that there is no detection from any of them, so that

$$P_d^c(t_0) = 1 - \left[1 - \Phi\left(\beta\sqrt{T} + \lambda\right)\right]^{k(t_0)}. \tag{5}$$

The probability of a detection from the first partially overlapping block $P_d^f(t_0)$ is given by

$$P_d^f(t_0) = \Phi\left(\frac{\beta(T - t_0)}{\sqrt{T}} + \lambda\right). \tag{6}$$

The probability of a detection from the last overlapping block $P_d^l(t_0)$ is given by

$$P_d^l(t_0) = \begin{cases} \Phi\left(\dfrac{\beta(\tau - T\lfloor \tau/T \rfloor + t_0)}{\sqrt{T}} + \lambda\right), & \text{if } 0 < t_0 < T(\lfloor \tau/T \rfloor + 1) - \tau \\ \Phi\left(\dfrac{\beta[\tau - T(\lfloor \tau/T \rfloor + 1) + t_0]}{\sqrt{T}} + \lambda\right), & \text{if } T(\lfloor \tau/T \rfloor + 1) - \tau < t_0 < T. \end{cases} \tag{7}$$

The overall probability of detection $P_d(t_0)$ is now given by

$$P_d(t_0) = P_d^c(t_0) + [1 - P_d^c(t_0)]\, P_d^f(t_0) + [1 - P_d^c(t_0)]\left[1 - P_d^f(t_0)\right] P_d^l(t_0). \tag{8}$$

The mean probability of detection $\overline{P}_d$, which is obtained by assuming that the transient signal can begin at any time during a block of samples or an integration period with equal probability, is given by

$$\overline{P}_d = \frac{1}{T} \sum_{m=0}^{T-1} P_d(m) \tag{9}$$

for the discrete-time case and

$$\overline{P}_d = \frac{1}{T} \int_0^T P_d(t)dt \tag{10}$$

for the continuous-time case. In the sequel, it will be shown that the sum in (9) for large $T$ can be replaced by an integral of type (10) with $T = 1$, and the resulting approximation is better for larger sample sizes $T$. Using the mean optimality criterion, the objective is to find the sample size or integration period $T^*$ which maximizes $\overline{P}_d$ according to the optimization problem

$$\begin{aligned} \textit{Maximize with respect to } T: \quad & \overline{P}_d \\ \textit{subject to:} \quad & T \in nbd(\tau) \\ & M_F \text{ fixed.} \end{aligned}$$

The sample size is restricted to an interval, denoted by $nbd(\tau)$, containing the point $T = \tau$. The endpoints of this interval are chosen so that the maximum of $\overline{P}_d$ with $T \in nbd(\tau)$ occurs in the interior of the interval and not at one of the endpoints. The reasons for adding this restriction were discussed in the introduction and will be clearer when the performance curves are discussed. No simplifications could be made to facilitate the computation of $T^*$ and we must resort to numerical techniques.

For the minimax optimality criterion, it is desired to find the sample size or integration period $T^{**}$ which maximizes the probability of detection for the least favorable transient signal arrival time. The optimization problem to find $T^{**}$ is

$$\begin{aligned} \textit{Maximize with respect to } T: \quad & \min_{0 \le t_0 < T} P_d(t_0) \\ \textit{subject to:} \quad & T \in nbd(\tau) \quad (11) \\ & M_F \text{ fixed.} \end{aligned}$$

Again, no simple analytical expressions could be found for $T^{**}$. However, it can be verified from (3) that $P_d(t_0)$ for $T > \tau$ and $t_0 \ge T - \tau$ has an axis of symmetry about the time $t_{00}$

given by

$$t_{00} = T - \tau/2.$$

For $T \leq \tau$, we find from (4)–(8) that $P_d(t_0)$ for $0 < t_0 < T(\lfloor \tau/T \rfloor + 1) - \tau$ has an axis of symmetry about the time $t_{01}$ given by

$$t_{01} = \frac{1}{2}\Big[T(\lfloor \tau/T \rfloor + 1) - \tau\Big],$$

and that for $T(\lfloor \tau/T \rfloor + 1) - \tau < t_0 < T$, $P_d(t_0)$ has an axis of symmetry about the time $t_{02}$ given by

$$t_{02} = \frac{1}{2}\Big[T(\lfloor \tau/T \rfloor + 2) - \tau\Big].$$

Thus, $\{t_{0i}\}$ must be the positions of local extrema of $P_d(t_0)$. Some preliminary numerical computations have indicated that one of these arrival times corresponds to a local and possibly global minimum for a variety of signal-to-noise ratios and values of the mean time between false alarms. Define the set $A(T)$ of arrival times by

$$A(T) = \begin{cases} \Big\{0^+, (T-\tau)^-, (T-\tau)^+, t_{00}, T^-\Big\}, & \text{if } T > \tau \\ \Big\{0^+, t_{01}, t_{02}, [T(\lfloor \tau/T \rfloor + 1) - \tau]^-, [T(\lfloor \tau/T \rfloor + 1) - \tau]^+, T^-\Big\}, & \text{if } T \leq \tau \end{cases}$$

which is the set containing the positions of the known local extrema, endpoints, and discontinuity points of $P_d(t_0)$. To simplify the numerical computation of $T^{**}$ for the minimax criterion in (11), the following optimization problem which finds an approximate solution $\hat{T}^{**}$ is substituted: the minimax criterion, (11) is replaced with

$$\begin{aligned} \textit{Maximize with respect to } T: \quad & \min_{t_0 \in A(T)} P_d(t_0) \\ \textit{subject to:} \quad & T \in nbd(\tau) \\ & M_F \text{ fixed.} \end{aligned}$$

Define $P_d^{wc}$ by

$$P_d^{wc} = \min_{t_0 \in A(T)} P_d(t_0). \tag{12}$$

## 5.3 Numerical Results

Using the formulas obtained in the previous section, curves were generated indicating the probability of detection versus the integration period for the continuous-time case. Figure 3 shows the mean probability of detection $\overline{P}_d$ as a function of the integration period for a transient of duration $\tau = 1$, a signal-to-noise ratio (SNR) of 20 dB, and a mean time between false alarms $M_F = 10^9$. The integration period is shown on a logarithmic scale so as to exhibit a rough symmetry about the local maximum occuring near $T = 0.58$. Note that the performance is arbitrarily good for large values of $M_F$ but the probability of false alarm $\alpha$ becomes unacceptable for these values. For example, for an integration period near $T = 10^9$, the probability of detection is near one but so is the probability of false alarm. For the reasons mentioned in the introduction, we will focus on the local minimum which occurs near $T = 1$ and thus the optimization is restricted to an interval about $T = 1$.

Another graph of the mean probability of detection versus the integration period is shown in Figure 4(a) for a transient of duration $\tau = 1$, an SNR of 5 dB, and a mean time between false alarms $M_F = 10^9$. Again, for large values of $M_F$, the performance is arbitrarily good. An enlargement of the region around $T = 1$ is shown in Figure 4(b) which indicates the presence of a local maximum. From this figure it is apparent that by choosing $T = 10^3$ the probability of detection will be greater than that obtained at the local maximum near $T = 1$. For $T = 10^3$, the probability of false alarm is $\alpha = 10^{-6}$ which might be quite acceptable. Nevertheless, we choose to focus on the local maximum near $T = 1$ for the reasons previously discussed.

Similar plots were generated for the worst case probability of detection $P_d^{wc}$ defined in (12). In Figure 5, the worst case probability of detection is plotted versus the integration

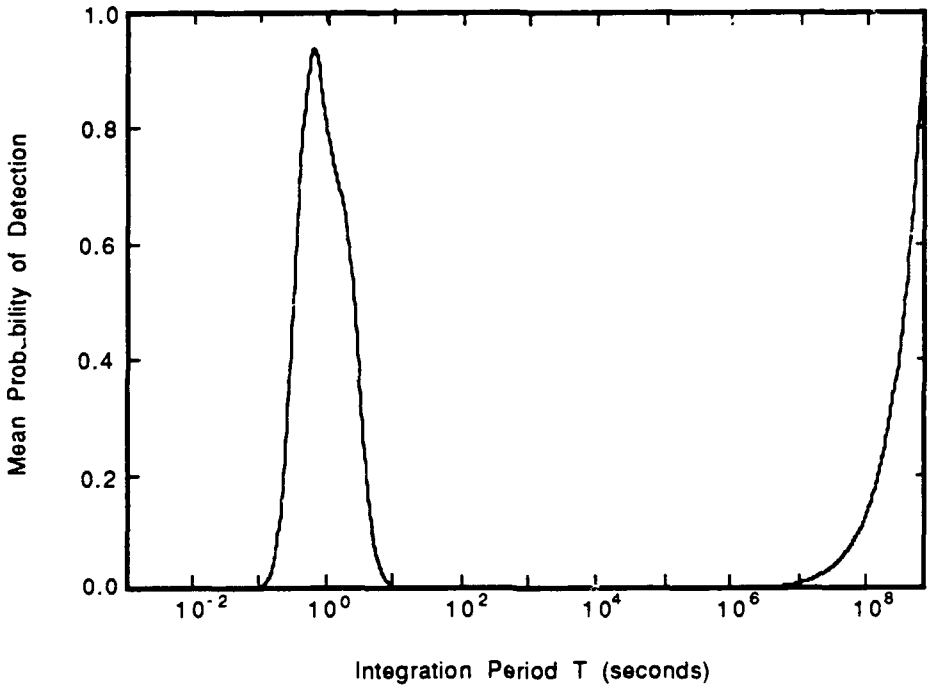
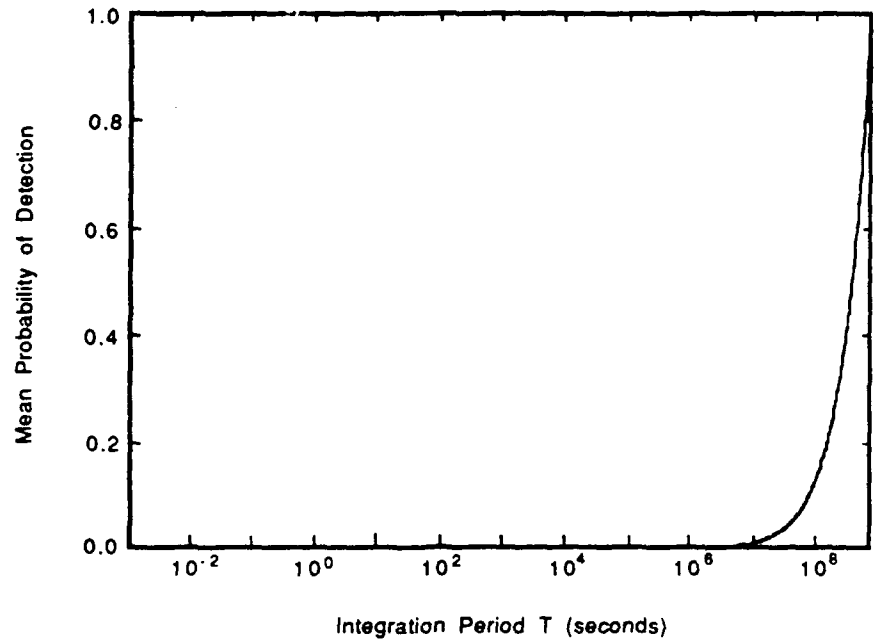
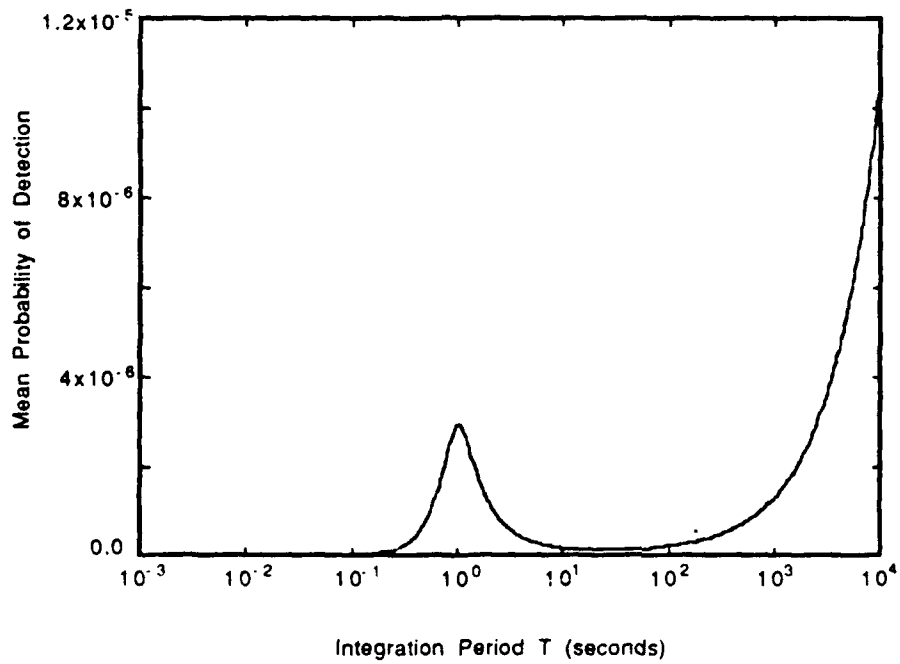



**Figure 3.** The mean probability of detection $\overline{P}_d$ as a function of the integration period $T$ for a transient duration $\tau = 1$, a signal-to-noise ratio of 20 dB, and a mean time between false alarms $M_F = 10^9$.




(a)




(b)

**Figure 4.** (a) The mean probability of detection $\overline{P}_d$ as a function of the integration period $T$ for a transient duration $\tau = 1$, a signal-to-noise ratio of 5 dB, and a mean time between false alarms $M_F = 10^9$. (b) An enlargement of the area around $T = 1$ showing the local maximum in the curve.

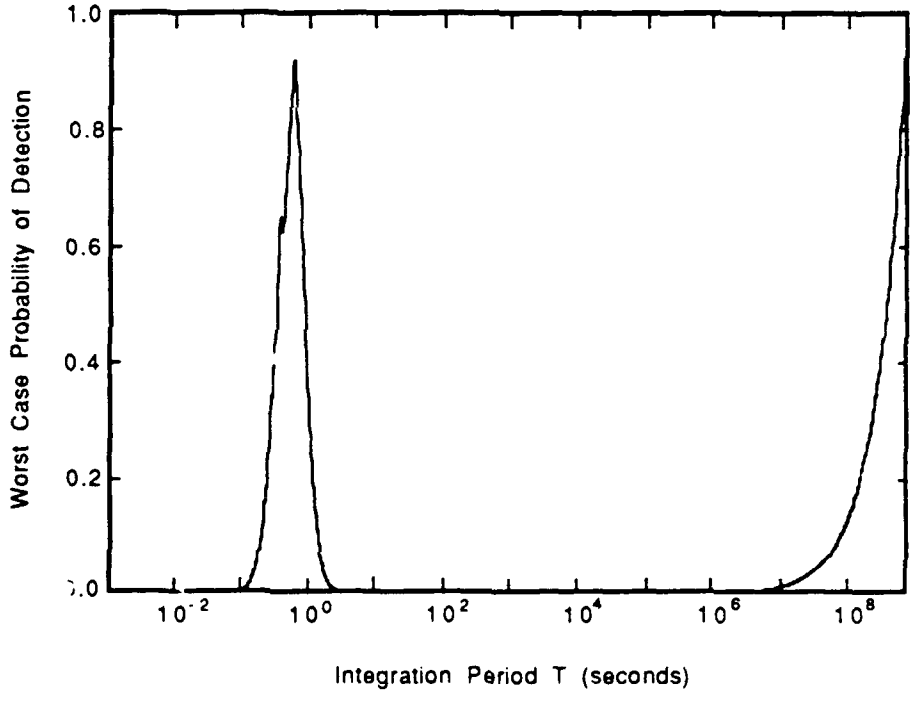
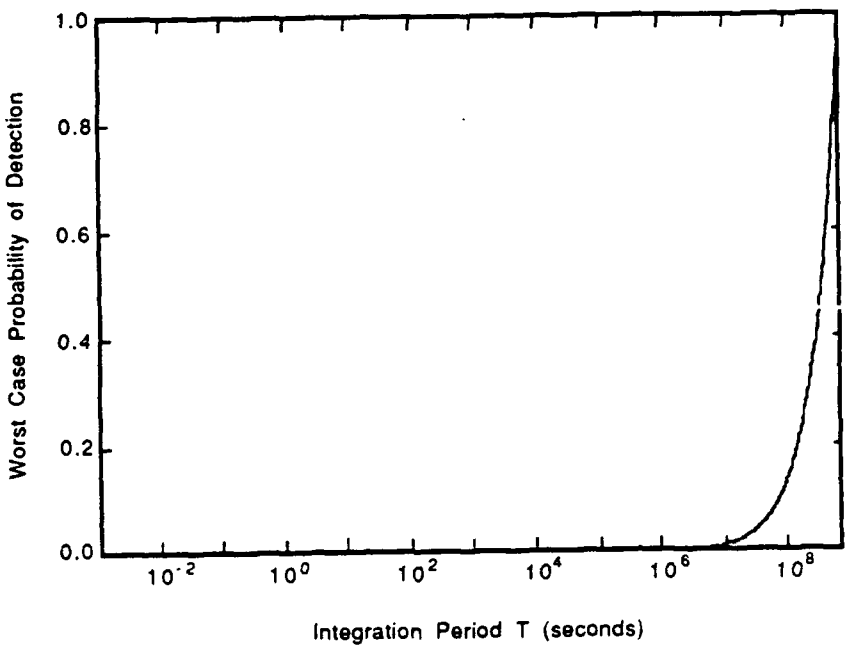
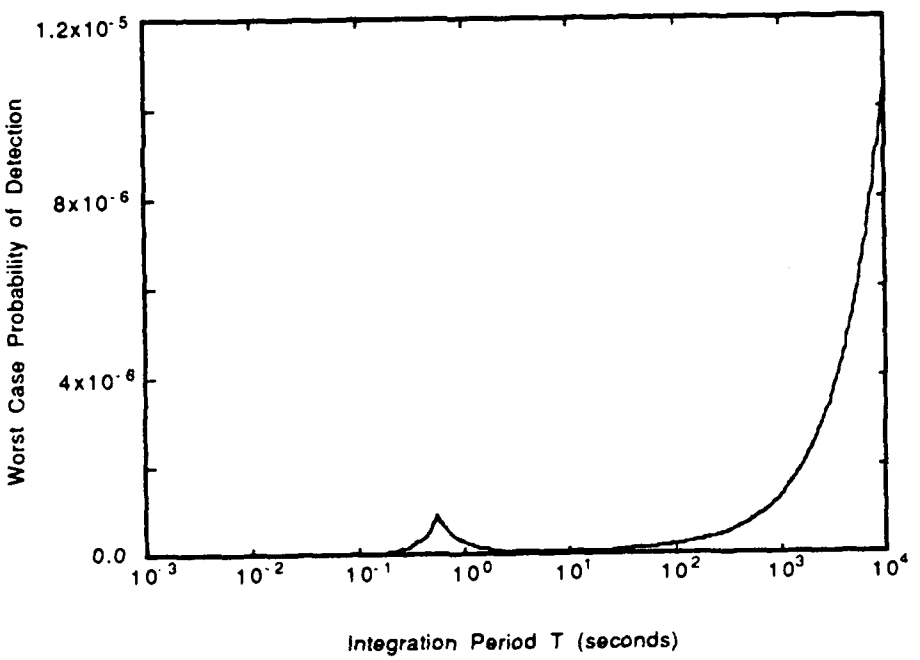



**Figure 5.** The worst case probability of detection $P_d^{wc}$ as a function of the integration period $T$ for a transient duration $\tau = 1$, a signal-to-noise ratio of 20 dB, and a mean time between false alarms $M_F \approx 10^9$.




(a)




(b)

**Figure 6.** (a) The worst case probability of detection $P_d^{wc}$ as a function of the integration period $T$ for a transient duration $\tau = 1$, a signal-to-noise ratio of 5 dB, and a mean time between false alarms $M_F = 10^9$. (b) An enlargement of the area around $T = 1$ showing the local maximum in the curve.

period with the same parameters as in Figure 3. The results for the parameters used in Figure 4 are shown in Figure 6. Again, the remarks made in connection with finding the optimal integration period to maximize the mean probability of detection apply to finding the optimal integration for the minimax criterion.

For a transient duration of $\tau = 1$, the optimal integration period was numerically computed as a function of the signal-to-noise ratio and the mean time between false alarms. It is shown in Figure 7 for the mean optimality criterion and in Figure 8 for the minimax optimality criterion. The optimal integration period for the minimax criterion varies very little with the SNR and the mean time between false alarms. The optimal value in this case is approximately $\hat{T}^{**} \approx 0.56$.

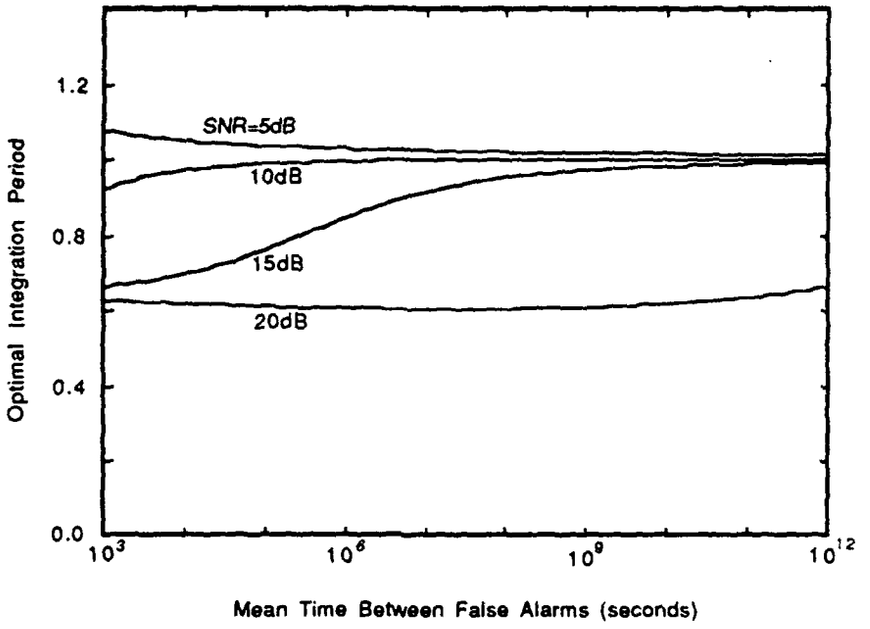
A procedure is now presented to determine the optimal integration period for either criterion for an arbitrary transient duration for the continuous-time case and for large sample sizes in the discrete-time case using the curves in Figures 7 and 8. To do this, it is useful to express the probability of detection $P_d(t_0)$ given in equations (3)–(8) as a function of all the relevant variables as follows:

$$P_d(t_0) = P_d(t_0, \tau, T, M_F, \beta, A, \sigma_0, \mu_0).$$

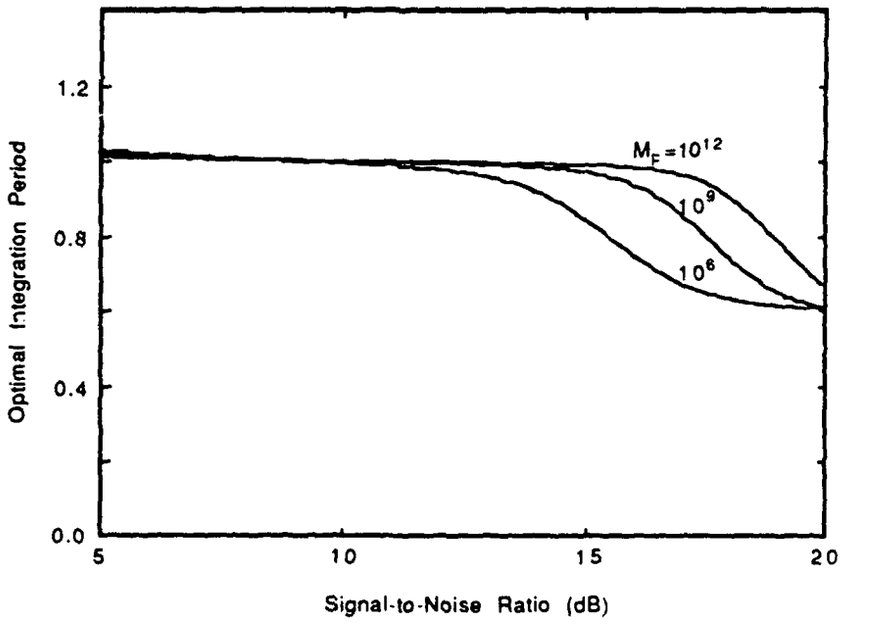
It is easily verified that

$$P_d(t_0, \tau, T, M_F, \beta, A, \sigma_0, \mu_0) = P_d\left(\frac{t_0}{\tau}, 1, \frac{T}{\tau}, \frac{M_F}{\tau}, \sqrt{\tau}\beta, \tau A, \sqrt{\tau}\sigma, \tau\mu_0\right).$$

Thus, the optimization with respect to $T$ for a given transient duration, mean time between false alarms, and SNR is equivalent to the optimization for a transient of duration $\tau = 1$ using the normalized parameter values.

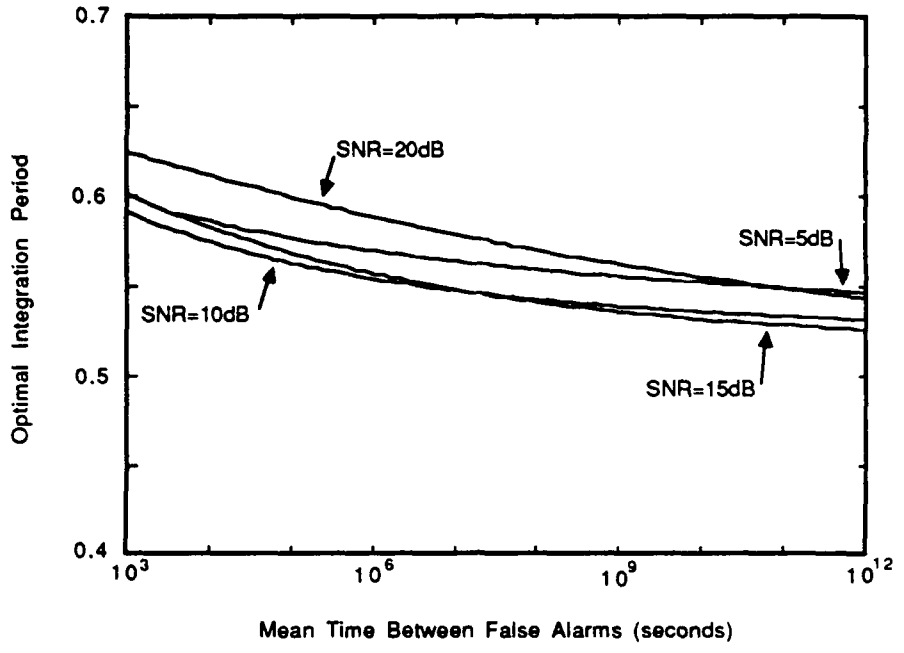



(a)

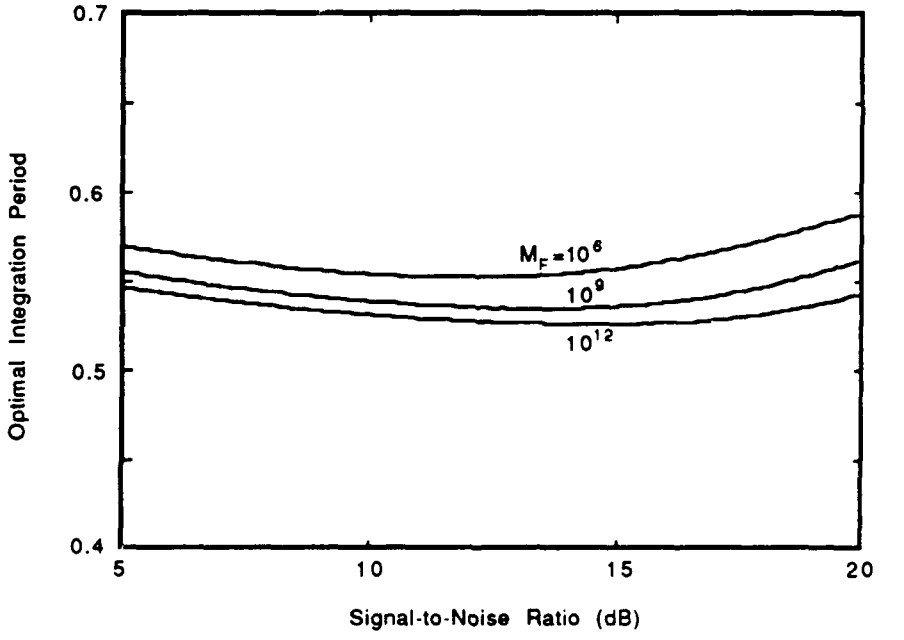



(b)

**Figure 7.** The optimal integration period $T^*$ for the mean criterion for a transient duration $\tau = 1$ (a) as function of the mean time between false alarms for a fixed signal-to-noise ratio and (b) as a function of the signal-to-noise ratio for a fixed mean time between false alarms.




(a)




(b)

**Figure 8.** The optimal integration period $\hat{T}^{**}$ for the minimax criterion for a transient duration $\tau = 1$ (a) as function of the mean time between false alarms for a fixed signal-to-noise ratio and (b) as a function of the signal-to-noise ratio for a fixed mean time between false alarms.

The procedure is as follows:

1. Find $T_1^*$ from Figure 7 or $\hat{T}_1^{**}$ from Figure 8 for a mean time between false alarms $M_F/\tau$ and an effective SNR of $\beta = \sqrt{\tau}A/\sigma$.

2. Compute $T^* = \tau T_1^*$ or $\hat{T}^{**} = \tau \hat{T}_1^{**}$.

3. Compute the threshold from (2), that is

$$h = \mu_0 T^* - \sqrt{T^* \sigma^2}\Phi^{-1}\left(\frac{T^*}{M_F}\right)$$

or

$$h = \mu_0 \hat{T}^{**} - \sqrt{\hat{T}^{**}\sigma^2}\Phi^{-1}\left(\frac{\hat{T}^{**}}{M_F}\right). \tag{13}$$

## 5.4. The Application to Narrowband Transients

In this section, the design procedure is applied to the problem of detecting a narrowband transient signal in white Gaussian noise. The problem formulation is almost exactly as set up in [2,3] for the quickest detection problem and it is repeated here with the appropriate modifications for the transient problem. The signal and noise are assumed to be real valued. The signal is a deterministic sinusoid with random phase. Length $N$ Discrete Fourier Transforms (DFT) are computed by the windowing and overlapping methods described in [1]. $N$ is chosen so that at the sampling frequency $f_s$, the frequency resolution $f_s/N$ equals the signal bandwidth. The single sample statistic $z_m$ is the squared magnitude of the coefficient corresponding to the frequency band of interest from the $m$'th DFT. Since the DFT's are overlapped, the statistics $\{z_m\}$ will be correlated, but the correlation will be nonzero only for a finite number of lags for white Gaussian noise. The random signal phase

is modeled as approximately constant for the duration of each DFT and independent of the phase during other DFT's as well independent of the noise. Let $M$ be the number of DFT's computed to form each test statistic $S_n$ which is to be optimized to detect a transient of length $K$ DFT's. It is assumed that $K$ and hence the optimal value of $M$ is large enough to ensure that test statistics have the required Gaussian distribution.

Proceeding as in [2,3], let $A$ be the signal amplitude, $\tilde{\omega}$ the frequency offset of the signal from center cell, $\sigma^2$ the noise variance, and $w(n)$, $n = 0, 1, \ldots, N-1$ the window coefficients. Let $r$ be the percent overlap of successive DFT's, i.e., $rN$ points overlap. From [1], we have

$$\mu_0 = \sigma^2 \sum_{n=0}^{N-1} w^2(n) = \sigma_0$$

$$\gamma_0 = \begin{cases} 1 + 2C^2(0.5), & \text{for 50\% overlap, } r = 0.5 \\ 1 + 2C^2(0.75) + 2C^2(0.5), & \text{for 75\% overlap, } r = 0.75 \end{cases}$$

where

$$C(r) = \frac{\sum_{n=0}^{rN-1} w(n)w(n + [1-r]N)}{\sum_{n=0}^{N-1} w^2(n)}.$$

The effective single sample SNR $\beta^2$ in terms of the input SNR $\beta_i^2$ in the signal bandwidth is given by [2,3]

$$\beta^2 = \beta_i^2 G(\tilde{\omega}) = \frac{NA^2}{4\sigma^2} G(\tilde{\omega})$$

where the gain $G(\tilde{\omega})$ is given by

$$G(\tilde{\omega}) = G_p \cdot [L_s(\omega)]^2 \cdot L_c$$

$$G_p = \frac{1}{N} \frac{\{\sum w(n)\}^2}{\sum w^2(n)}$$

$$L_s(\omega) = \frac{\left|\sum w(n) e^{-j\omega n}\right|}{|\sum w(n)|}$$

$$L_c = \frac{1}{\sqrt{\gamma_0}}.$$

$G_p$ is the window processing gain or the reciprocal of the equivalent noise bandwidth, $L_s(\omega)$ is the scalloping loss, and $L_c$ is the loss due to the correlation between overlapped DFT's.

**Example.** Suppose we are trying to detect a transient of length $K = 100$ DFT's with an input SNR per cell of $-5$ dB. A Hamming window is used with 50% overlap. We assume the worst case scalloping loss so that $\tilde{\omega} = \pi/N$. From [1, table 1], we have that [2,3]

$$10\log_{10} G = -3.35 \text{ dB},$$

so that the effective SNR is

$$\beta = -8.35 \text{ dB}.$$

We will use the minimax optimality criterion to find the optimal number of DFT's $M^*$ for a mean time between false alarms $M_F = 10^8$ DFT's. To use Figure 8, the normalized SNR and mean time between false alarms are computed by

$$\beta' = \beta + 10\log_{10}(K) = -8.35 + 20 = 11.65 \text{ dB}$$

$$M_F' = M_F/100 = 10^6 \text{ DFTs}.$$

From Figure 8(b), $\widehat{T}^{**} = 0.55$ so that $M^* = (0.55)(100) = 55$ DFT's. From [1,2,3], $\mu_0 = \sigma_0 = 0.397N\sigma^2$ and $\gamma_0 = 1.11$. The threshold $h^*$ is computed from (13),

$$\begin{aligned} h^* &= \mu_0 M^* - \sqrt{M^*\gamma_0\sigma_0^2}\Phi^{-1}\left(\frac{M^*}{M_F}\right) \\ &= \mu_0\left[M^* - \sqrt{M^*\gamma_0}\Phi^{-1}\left(\frac{M^*}{M_F}\right)\right] \\ &= (36.95)N\sigma^2. \end{aligned}$$

## 5.5 Conclusion

In this chapter, the problem of the optimal sample size or integration period for the detection of a transient signal was examined. While the framework for the discrete-time problem was more general than that for continuous time, both cases relied on using the Gaussian distribution for the test statistics. For the continuous-time situation, a set of

curves was numerically computed which indicate the optimal integration period for both the mean optimality criterion and the minimax optimality criterion. A design procedure was given to use these curves for transients of arbitrary length in the continuous-time case and transients consisting of a large number of samples in the discrete-time case. It was shown that the results could be applied to the detection of narrowband transient signals.

Throughout this chapter, it has been assumed that the duration of the transient to be detected is known to a high degree of accuracy. If the mechanics of the phenomenon which generates the transient is well understood, then this duration may be known. However, in some situations only an estimate is available or the duration may in fact be a random variable. Now, the results presented in this chapter are no longer optimal although one may still want to use the "optimal" sample size or integration period for the average transient duration. In such a situation, this work serves more as a general guide in the design of the detector. It demonstrates that the integration period should be one half to one times the true transient duration.

## 5.6 References


[1] F. J. Harris, "On the use of windows for harmonic analysis with the Discrete Fourier Transform," *Proc. of the IEEE*, vol. 66, no. 1, pp. 51–83, January 1978.

[2] L. Pelkowitz and S. C. Schwartz, "Asymptotically optimum sample size for quickest detection," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-23, no. 2, pp. 263–272, March 1987.

[3] L. Pelkowitz, *The General Disorder Problem in Discrete Time*, Ph.D. Dissertation, Department of Electrical Engineering, Princeton University, Princeton, N.J., June 1987, Chapter 5.

[4] A. N. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process," *Sov. Math. Dokl.*, no. 2, pp. 795–799, 1961.

# — 6 —

# Conclusion

## 6.1 Summary and Future Work

This dissertation investigated sequential decision procedures to detect changes in the statistical model of an observed random process when these changes can occur at unknown times. Specifically, the disorder problem and the transient problem were examined with the appropriate performance criteria for the application. Below, the results of this study are summarized along with suggestions for further research.

In Chapter 2, Page's test was generalized by considering arbitrary nonlinearities for detecting a change in distribution. A simple performance measure was defined which corresponds to the slope of the performance curves for a large mean time between false alarms. An analytic bound was found for this measure which allows the performance of different nonlinearities to be easily compared. Also, by considering the local performance of Page's test, it was shown that the measure coincides with classical detector efficacy in the ordinary binary hypothesis testing situation. These results relied on the assumption that the observations before and after the disorder were independent and identically distributed

random variables. While there have been some results for the quickest detection problem with dependent observations using a likelihood ratio approach, there has been no work on using variants of Page's test with general nonlinearities for the dependent case. The recent results on extending Wald's fundamental identity to certain dependent sequences [3] might provide the necessary framework for the development of a performance measure for Page's test using dependent observations paralleling the work in Chapter 2.

Chapter 3 considered the application of Page's test to the detection of transient signals. The test statistics were derived from the DFT of the observations to allow for the detection of a change in the energy spectrum. Several other sequential tests were also considered for this problem. Using Monte Carlo simulations, the performances were compared and it was found that Page's test performed as well as or better than the other tests in nearly all situations. This work represents the first time that the performances of these sequential tests are evaluated in the context of transient detection. There is much more work that could be done on the analytic evaluation of sequential tests. Theoretical questions such as the expected hitting times of moving average processes could be investigated to provide a more complete framework for analyzing sequential tests.

The Gabor representation was investigated in Chapter 4. It was shown that such a representation could be useful in detecting transient signals in noise. It was further demonstrated that for the purposes of detection and estimation, the maximum likelihood estimates of the Gabor coefficients may be more appropriate to use than those obtained through the biorthogonal function procedure. Several issues remain unresolved, namely, the effects of window mismatch, delay mismatch, and frequency mismatch. Moreover, when mismatch is expected, a window function which is robust over an uncertainty class of possible signals might be desirable, similar to the idea of robust matched filtering. Also, the merits

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

of this approach should be compared to more traditional methods such as the short time Fourier Transform when the representation does not exactly match the actual situation. One might also want to investigate the relative merits of the Gabor representation and the more traditional windowing methods for the DFT used to enhance resolution [1]. Finally, the wavelet representation [2] has recently received much attention as an alternative to the Gabor representation but no detection schemes have been formulated and evaluated using wavelets.

The optimal sample size for a transient detector was explored in Chapter 5. A procedure was given to determine the optimal sample size or integration period for the detection of a simple transient with known duration. Further work could be done on robustifying the detection procedure to allow for arbitrary transient lengths. Perhaps this could be accomplished by using detection statistics corresponding to different time resolutions as might result from utilizing the wavelet representation.

As a whole, this dissertation has focused on sequential detection procedures and their evaluation with appropriate performance criteria. Since many real signal processing schemes are implemented in a sequential fashion but are usually only evaluated based on fixed sample size performance, this is an area that warrants further study.

## 6.2 References


[1] F. J. Harris, "On the use of windows for harmonic analysis with the Discrete Fourier Transform," *Proc. of the IEEE*, vol. 66, no. 1, pp. 51–83, January 1978.

[2] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[3] J. S. Sadowsky, "A dependent data extension of Wald's identity and its applications to sequential test performance computation," *IEEE Trans. Info. Theory*, vol. 35, no. 4, pp. 834–842, July 1989.